Elucidation of the sequence selective binding mode of the DNA minor groove binder adozelesin, by high-field 1H NMR and restrained molecular dynamics

Cameron, Linda

Award date:
1999

Awarding institution:
University of Bath

Elucidation of the Sequence Selective Binding Mode of the DNA Minor Groove Binder Adozelesin, by High-Field ^{1}H NMR and Restrained Molecular Dynamics.

Submitted by

Linda Cameron

for the degree of PhD

of the University of Bath

1999

The research work carried out in this thesis has been carried out in the Department of Pharmacy and Pharmacology, under the supervision of Dr. Andrew S. Thompson and Dr. Michael D. Threadgill.

L. Cameron

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

Adozelesin (formerly U73-975, The Upjohn Co.) is a covalent, minor-groove binding analogue of the antitumour antibiotic (+)CC-1065. Adozelesin consists of a cyclopropa-pyrroloindole alkylating sub-unit identical to (+)CC-1065, plus indole and benzofuran sub-units which replace the more complex pyrroloindole B and C sub-units, respectively, of (+)CC-1065. Adozelesin is a clinically important drug candidate, since it does not contain the ethylene bridge moieties on the B and C sub-units which are thought to be responsible for the unusual delayed hepatotoxicity exhibited by (+)CC-1065. Sequencing techniques identified two consensus sequences for adozelesin binding as p(dA) and 5′(T/A)(T/A)T-A*(C/G)G. This suggests that adozelesin spans a total of five base-pairs and shows a preference for A=T base-pair rich sequences, thus avoiding steric crowding around the exocyclic NH_2 of guanine and a wide minor groove. In this project, the covalent modification of two DNA sequences, *i.e.* 5′-d(CGTAAGCGCTTA*CG)$_2$ and 5′-d(CGAAAAA*CGG)· 5′-d(CCGTTTTTCG), by adozelesin was examined by high-field NMR and restrained molecular mechanics and dynamics. Previous studies of minor groove binding drugs, using techniques as diverse as NMR, X-ray crystallography and molecular modelling, indicate that the incorporation of a guanine into the consensus sequence sterically hinders binding and, more importantly, produces a wider minor groove which is a 'slack' fit for the ligand.

The aim of this investigation was to provide an insight into the sequence selective binding of adozelesin to 5′-AAAAA*CG and 5′-GCTTA*CG. The ^{1}H NMR data revealed that, in both cases, β-helical structure and Watson-Crick base-pairing was maintained on adduct formation. The 5′-GCTTA*CG adduct displayed significant distortion of the guanine base on the non-covalently modified strand. This distortion resulted from an amalgamation of two factors. Firstly, the presence of a strong hydrogen-bond between the amide linker of the indole and benzofuran sub-units and the carbonyl of a central thymine base plus a second, weaker, hydrogen-bond to the exocyclic NH_2 of the guanine. Secondly, covalent modification of an adenine by adozelesin yields a distortional wave which terminates 6 base-pairs upstream in the 5′-direction. ^{1}H NMR data and the resulting molecular model show that adozelesin lies closer to the modified strand as a consequence of this strong hydrogen-bond and perhaps also of the distorted guanine pushing adozelesin to the opposite side of the groove. No distortion of this kind was noted in the 'A' tract model; however, the ^{1}H NMR data shows that the adenine bases remain propellor twisted, which is characteristic of macroscopically bent DNA. This molecular model showed that, within this most preferred sequence, adozelesin lay centrally within this straighter 'A' tract, yielding an improved complementary fit of the drug within the DNA minor groove. No hydrogen-bonds were detected. This study indicates the possible role of the ureylene linker in the selective binding of adozelesin to the unpredicted GC-containing consensus

sequence. Various other minor groove binders (cyanomorpholinoadriamycin, DSB-120 and sibiromycin) also had their sequence-selectivity properties probed with limited success, owing to their inability to yield a single drug-DNA adduct for high-field NMR study.

The adozelesin-DNA self-complementary adduct in this study has provided a new and exciting starting point for rational drug design of an adozelesin dimer, a proposed antineoplastic agent with enchanced sequence-selectivity. These adozelesin-DNA adducts have proved to be an excellent model system to investigate the molecular basis for DNA recognition and binding.

Acknowledgements

There are many people I would like to thank for their help and support during my postgraduate studies. Firstly, I would like to thank Dr. Andrew S. Thompson for providing me with the opportunity to study for a PhD at the University of Bath, Department of Pharmacy. Dr. Michael D. Threadgill for proof-reading this thesis and his many thoughtful comments and suggestions, which were greatly appreciated. Dr. George Gray at Varian, Palo Alto for the accumulation of the 600 MHz NMR data. Mr. R. H. Hartell and Mr. D. Wood for providing some of the one-dimensional NMR experimental data. Dr. R. Kinsman, Department of Chemistry at the University of Bath, for his help and assistance in the accumulation of the 400 MHz two-dimensional NMR data. Mr. Kevin C. Smith for his friendship and support during my time in Bath and for providing me with new perspectives on many computer modelling problems when I began to get tunnel vision.

I would also like to say a huge thankyou to all my family and friends for their encouragement in times of despair, of which there were many. I would like to thank my flatmate Louise Kelly for her friendship and support. Mario Buglione for his endless encouragement and patience while I completed this project. My Mum, for believing in me and making me persevere during this last few years.

CONTENTS

FIGURE CONTENTS

Abbreviations

A	Adenine
G	Guanine
T	Thymine
C	Cytosine
I	Inosine
NMR	Nuclear Magnetic Resonance
NOE	nuclear Overhauser effect
p	poly *i.e.* poly adenine is p(A)
d	deoxyribonucleotide
HPLC	high pressure liquid chromatography
2D	two-dimensional
1D	one-dimensional
DNA	deoxyribonucleic acid
J	coupling constant
ppm	part per million
Hz	hertz
NOESY	nuclear Overhauser effect spectroscopy
ROESY	rotating frame Overhauser spectroscopy
TOCSY	total correlation spectroscopy
COSY	correlation spectroscopy
DQF	double quantum filter
CPI	cyclopropapyrroloindole
RNA	Ribonucleic acid
Ado	adozelesin
EDTA	Ethylenediaminetetra-acetic acid
CMA	cyanomorpholinoadriamycin
Pu	purine
Py	pyrimidine
FID	Free induction decay
NAD	Nicotinamide adenine dinucleotide
FT	Fourier transform
D1	dimension one
D2	dimension two
t_1	evolution time
F_1	frequency domain
F_2	frequency domain
δ	chemical shift

τ	mixing time
Å	angstrom
PBD	pyrrolo[2,1-*c*][1,4]benzodiazepine
k_b	binding affinity
cov	covalent
noncov	non-covalent
'A' tract	adenine rich oligonucleotide
'T' tract	thymine rich oligonucleotide
5′	polarity directed to 5′ phosphate terminal of DNA
3′	polarity directed to 3′ terminal of DNA
i	intensity
γ	magnetogyric ratio
r	distance
E	energy
W	probability of single quantum transition
PPSS	peak pick spreadsheet
MSS	master spreadsheet
Kr	rate of reaction
Im	imidazole
Py	pyrrole
Dp	N,N-dimethylaminopropylamide
Gamma	gamma-aminobutyric acid
Ac	acetyl
Beta	p-alanine (hairpin polyamides)

CHAPTER 1: INTRODUCTION

Half a century has passed since the first tentative uses of the nitrogen mustards in the treatment of cancer. Today, even in the era of modern high technology, after billions of pounds have been spent and a truly colossal effort into research and development has been made, there is still no 'wonder drug' to cure cancer. Compounds such as adriamycin and cisplatin are clinically effective against a broad spectrum of tumours but some neoplasias still exhibit little response. All anticancer agents have the disadvantage of unpleasant side effects, though it is hoped in most cases that these will be more than compensated for in quality and quantity of life.

Drug design can be defined as, 'the approaches and steps which are involved in the production of new chemical entities of pharmacological importance'. From a pharmacodynamics standpoint, the ideal cancer drug should be: receptor selective, with extremely high potency and largely free of side effects. Whilst these factors are fundamental to new drug development, from a commercial stance also, the reality is that anticancer treatments address a sizeable clinical need and the state of medical knowledge is such that it justifies the funding risk.

There are four major sources of new drugs. Some old drugs may evolve to have a new use, possibly originating from their side-effects. Natural products which are not found in humans may have activity as a drug, for example; plant compounds, antibiotics from bacteria or fungi, or products from marine animals *e.g.* bryostatin. Chemicals which are found in humans that exhibit activity as a drug, like hormones *e.g.* insulin, vitamins and peptides *e.g.* enkephalins. Lastly there are novel chemical syntheses, which can proceed either randomly, accidentally or rationally when the receptor, agonist or antagonist structure is known. New anticancer agent development which focuses mainly on the synthesis of analogues, by modification of functional groups to enhance potency, pharmacokinetics and selectivity, has been employed extensively. The prodrug carzelesin, which is an analogue of the natural product (+)CC-1065, is an example of this approach (Li *et al.*, 1992).

Today, it is becoming more popular to examine the target, for example topoisomerases I and II, or look for a new one. Techniques such as high-field NMR, X-ray crystallography, computer molecular modelling and the use of molecular biology for DNA sequencing and gene cloning are readily available. They provide a wealth of information on receptor-substrate topology, mechanism of interaction and the resulting modulation of effects, both at the molecular and whole system level. Other novel approaches to improve the selectivity of these cytotoxic drugs include the use of immunoconjugates or polymer conjugates (Verweij, 1996).

The cytotoxic drug (*e.g.* adriamycin) is attached to a tumour-specific antibody (*e.g.* BR96) which will then be readily taken up by the cancer cell, carrying the drug directly to the site of action (Onetto *et al.*, 1995). Polymers may act in a similar manner (Duncan, 1992), making use of a unique cancer cell enzyme to cleave for example, a peptide linkage which bridges the ligand and the polymer, releasing the ligand in its active form. Ideas such as these could be of key importance for drug targeting.

DNA is known to be the intracellular target of many cancer chemotherapy agents. Their mode of action is *via* blockage of DNA or RNA synthesis, perhaps by prevention of an enzyme binding at a specific sequence or progressing along DNA. The enormity of the sequence selectivity problem arises with the need to target only a tiny consensus sequence, a few base pairs long, within the $2{\cdot}9x10^9$ base human genome. This is comprised of only adenine, guanine, thymine and cytosine in a large number of permutations. Identification of longer consensus sequences, which would probably occur less frequently within cellular DNA, could be a solution to the problem of lack of specificity. The design of chemotherapeutic agents which could span larger numbers of bases would then be required. However, as becomes clear in the following review of DNA-binding drugs, characterization of smaller segments of DNA and elucidation of the mechanisms by which naturally occurring antitumour antibiotics interact with them, constitutes a myriad of problems in itself. This is prior to attempts to design and synthesize superior analogues, which sometimes seems to be akin to molecular roulette!

There are three possible sites on DNA to which a molecule can bind: the major groove, the minor groove or between the bases. The major groove is characteristically deep but very wide. It is lined with the potential hydrogen-bond acceptor atoms; N-7 of guanine and adenine (*n*), O-4 of thymine (*o*) and O-6 of guanine (*o*), while the amino group at the C-6 of adenine and C-4 of cytosine act as hydrogen bond donors (*h*) (Stryer, 1988). Therefore the patterns of donor and acceptor atoms are as follows: (AT) *nho*, (TA) *ohn*, (GC) *noh* and (CG) *hon*. It is this mixture of a large number of hydrogen-bond interactions and width induced accessibility, which renders the major groove the site of choice to where large proteins (*e.g.* transcription factors and enzymes) can interact sequence-selectively with DNA. The depth is conducive to small drug molecules binding; however, the width makes this site unfavourable to many small ligands, especially if they are hydrophobic in nature. Sequence selectivity is difficult to achieve in an environment where the drug (key) is a slack fit in the major groove receptor (lock), in effect allowing a molecule of any shape to bind at any position. Intercalation of drug betwen the bases provides another binding mode *e.g.* adriamycin (Capranico *et al.*, 1986). The problem with intercalation is that the sequence specificity only extends to the pair of base pairs which encase the drug *e.g.* d(CG) or d(GC) for adriamycin (Chaires *et al.*, 1987). There will be a large number of CG or GC base sequences, leading to a relatively non-specific

binding mode. Non-specificity is very undesirable because it results not only in a chemotherapeutic action at the target sequence but also in reaction at many other sites giving rise to mutations or side-effects.

The minor groove, however, has many qualities which promote it as the intracellular drug target of DNA. It is deep and narrow in β-helical DNA, so that a drug binding within it will be able to fit tightly within the hydrophobic environment of the minor groove. The large surface area of the groove provides opportunities not only for stabilizing hydrogen-bonds, but also for strong forces of attraction generated from many van der Waals forces. Bases within the DNA can be alkylated to form a covalent adduct. The N-3 of adenine and guanine, and O-2 of thymine and cytosine constitute hydrogen-bond acceptor atoms of the minor groove, whilst the amino group of guanine is the sole hydrogen donor. Hence, the minor groove has less hydrogen-bonding possibilities than the major groove *i.e.* (AT) *no*, (TA) *on*, (GC) *nho* and (CG) *ohn*. It would therefore follow that minor groove binding would be less sequence specific, due to the lack of distinction between AT or TA base-pairs (both are hydrogen-bond acceptor, acceptor) and CG or GC base-pairs (acceptor, donor, acceptor). This equivalence of base-pairing is not observed within the major groove, explaining the high fidelity by which proteins can read the DNA sequence. However, it is the topology of the minor groove which is the most important contributing factor to the achievement of sequence selectivity within the minor groove: 'A' tracts give rise to extremely narrow grooves for a close complementary fit with the ligand; GC regions can accommodate bulkier molecules. The most important reason for designing drugs which target DNA is because it is at the core of every cell's ability to sustain life. Therefore, targeting the minor groove as a drug receptor could lead to modulation of processes such as DNA replication to achieve a therapeutic effect.

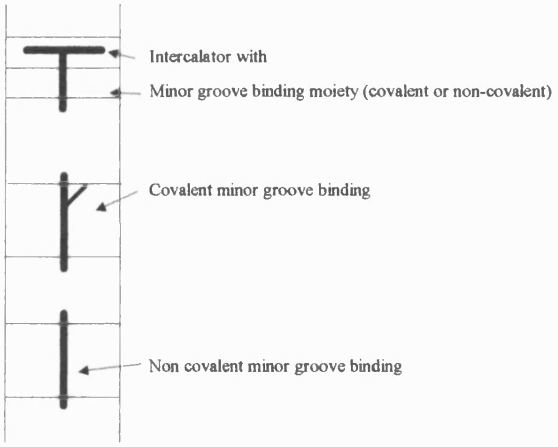


Figure 1 Schematic diagram of how the various families of ligands bind to DNA.

There are three possible modes by which ligands interact with the minor groove of DNA (Fig. 1). These are;

1) Noncovalent minor groove binding
2) Covalent minor groove binding
3) Minor groove binding with intercalation

1.1 Noncovalent minor groove binders

There are four major classes of non-covalent minor groove binders: the oligopeptide antibiotics, the bis-benzimidazoles, the bisamidines and the bis(quaternary ammonium) heterocycles. Their modes of binding and specificity will now be reviewed in detail.

1.1.1 Oligopeptide antibiotics

Netropsin and distamycin A are both naturally occurring peptide antibiotics. The former was isolated from *Streptomyces netropsis* (Finlay *et al.*, 1951) and displays a range of antiviral, antitumour and antibacterial activity against both gram-positive and gram-negative bacteria (Hahn, 1975; Zimmer, 1975). Distamycin A similarly exhibits these types of biological activities and inhibitory effects. Neither are in general clinical use, except as topical applications. Distamycin A was discovered by Arcamone *et al.*, in 1964, produced by *Streptomyces distallicus*. Both netropsin and distamycin A belong to a class of compounds known as pyrrole-amidine antibiotics and are di- and tri- peptides respectively. The structure of netropsin (Fig. 2) consists of two N-methylpyrrole rings with an amide linkage between them, and either side of them. A guanidium group is at the amino terminus and a propylamidinium tail is attached to the carboxyl terminus. Distamycin differs only in having three N-methylpyrrole carboxamide subunits and a formamide instead of the guanidium head unit (Fig. 2) (Kopka & Larsen, 1992).

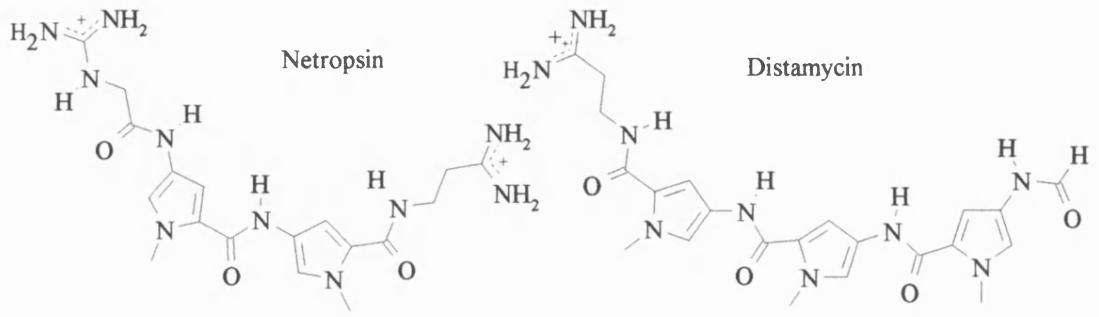


Figure 2 The molecular structures of netropsin and distamycin.

The biological activities and sequence selectivities of these agents on binding to DNA have been investigated in great detail. X-ray crystal structures of netropsin and distamycin (Berman *et al.*, 1979; Gurskaya *et al.*, 1979) indicate that the rings and, hence, the backbone of these drugs are crescent-shaped, with the concave side containing the amide groups. This structural feature makes these ligands perfect complementary fits for insertion into the minor groove of β-DNA.

Zimmer *et al.*, (1975, 1980) and Wartell *et al.*, (1974) have shown that netropsin has a high preference for binding to the DNA minor groove, when it is rich in adenine and thymine base pairs. Wartell *et al.*, (1974) proposed that the positively charged ends of netropsin, interact with the 5′-phosphate anionic oxygens either side of an AT base pair, so the ligand spanned the width of the minor groove. A later binding model (Berman et al., 1979) placed netropsin lengthwise across four base pairs within the groove. Hydrogen-bond details were included between amide nitrogen atoms and thymine O2 (TO2) or adenine N3 (AN3) for stabilization of the complex. The structures of netropsin-DNA complexes from later studies are similar to this proposed model.

Patel (1982) used NMR techniques to probe the binding of netropsin to 5′-$d(C^1G^2C^3G^4A^5A^6T^7T^8C^9G^{10}C^{11}G^{12})_2$ in solution. DNA oligomers are numbered sequentially from 1 to n in the direction of 5′- to 3′-. The complementary strand is numbered in the same manner beginning at n+1 if the sequence is non-self-complementary, or if the sequence is self-complementary the numbering is identical to the first strand *i.e.* 1 to n. Intermolecular nuclear Overhauser effect (NOE) measurements were used to constrain a model of the complex, with the pyrrole H3 protons positioned adjacent to A^6H2 and the guanidino methylene protons next to A^5H2. Netropsin spans four base pairs in total. Hydrogen-bonds, as Berman *et al.*, (1979) also described, occur between AN3 and TO2 of DNA and the three amide peptide groups of netropsin. Patel (1982) used the spatial confinements of an AT rich minor groove to explain the sequence selectivity of this ligand. Owing to guanine having a bulky amino group which protrudes into the minor groove, netropsin would be obstructed from binding and forming the above described stabilizing interactions.

This non-intercalative DNA-binding drug has also been studied by Zakrzewska *et al.*, (1983) using a computational approach. This comparative study looked at the interaction of SN18071, a bisquaternary heterocycle and netropsin with a poly(dA)·poly(dT) DNA sequence. It was concluded that hydrogen-bonds increase sequence-specificity but are not a prerequisite for binding to the oligomer, due to the ability of SN18071 to conform to the spatial restrictions of the AT minor groove, even though it has no hydrogen-bonding capacity. The minimum energy for the netropsin-poly(dA)·poly(dT) complex was lower than that generated with SN18071. This demonstrated the preference of netropsin for an AT site, which has a low

electrostatic potential and, hence, attracts the positively charged drug. As a consequence, the fit of netropsin into the minor groove is far superior to that of SN18071.

X-ray diffraction data were collected on the netropsin-5′-d(CGCGATATCGCG) complex (Coll *et al.*, 1989) to examine why the binding affinity of netropsin was greater for poly(A)·poly(T) sites rather than poly(AT). Binding took place at the AT-rich segment of the minor groove, with netropsin lying in both orientations. Instead of the usual three-centred bifurcating hydrogen-bonding as described by Berman *et al.*, (1979), only single H-bonds were present due to alternating AT bases, giving rise to short and long H-bond distances from TO2 or AN3 to netropsin NH groups. Netropsin bound to the 5′-ATAT groove, which is slightly wider than a homologous A strand duplex, owing to the lack of resulting base-pair propellor twist. The sugar-phosphate backbone was pulled towards netropsin inwards *via* van der Waals forces to create a narrower groove which is a snugger fit for netropsin. This conformational rearrangement, in effect, creates the same environment as the poly(A)·poly(T) duplex, which is the preferred binding site of netropsin, probably due to this extra incurred energy penalty on poly(AT) oligomers.

Distamycin-DNA interactions bear a close resemblance to those of netropsin. ^{1}H NMR (nuclear magnetic resonance) has been used to investigate the interaction of distamycin A with 5′-d(CGCGAATTCGCG)$_2$ (Klevit *et al.*, 1986), which also binds within the minor groove. Very intense NOEs were noted between the pyrrole H3s of rings one, two and three and the adenine H2s of A^5, A^6 and A^{18}, respectively. This work also demonstrated that the 5′-d(AATT)3′ base pair sequence was large enough to bind distamycin, without requiring another AT in the binding site as Kopka *et al.*, 1985, and Coll *et al.*, 1987 suggested. The latter reported that, when distamycin binds to 5′-d(CGCAAATTTGCG)$_2$3′, the ligand covered five of the six bases. Klevit *et al.*, (1986) stated that there was little difference in the binding affinity of distamycin to four or five successive AT base pairs. This indicated that while only four AT base pairs may be required for selectivity, the extra length (due to the third pyrrole) required that this ligand extend into a fifth base pair region. The orientation of the drug is dependent on the DNA flanking sequences around the binding site (Dervan *et al.*, 1986).

Pelton *et al.*, (1988) tried to probe further the important factors for the interaction of distamycin A with 5′-d($C^1G^2C^3G^4A^5A^6T^7T^8C^9G^{10}C^{11}G^{12}$)$_2$ using 2D NMR techniques and molecular modelling. By analogy with netropsin, distamycin formed three-centred, or 'bifurcating' hydrogen-bonds between the amide group of distamycin and the adenine N3, thymine 02 atoms within the minor groove (Fig. 3). Van der Waals contacts relating to sequence-selectivity were also seen as previously described, from adenine H2 to pyrrole H3 protons. It is thought that a mixture of hydrogen-bonds and electrostatic forces are responsible

for the strong binding of distamycin. DNA O′1 atoms were stacked over each pyrrole ring, the third ring rotating out of plane slightly, in order to follow the same pitch as the minor groove.

Pelton *et al.*, (1989; 1990), titrated distamycin A with oligomers of expanding AT target sequences and examined them by NMR. At low drug concentrations *i.e.* less than a 1:1 drug:DNA stoichiometry, two species of distamycin A bound to DNA complex existed. A mixture of these oligomers bound to one and two distamycin A molecules simultaneously, head to tail alongside each other (so that the +ve charges were not adjacent to each other) were encountered. Higher drug:DNA concentration ratios resulted in only the 2:1 complex being present. The fact that only one antiparallel complex was accounted for means distamycin exhibits the ability to read the minor groove sequence selectively, which is why it is known as a 'lexitropsin'. Binding is similar to that described previously, minus the bifurcating H-bonds. This is due to the spatial confinements of having two distamycin molecules in the groove, each having to lie alongside a wall instead of within the centre of the minor groove. Because one ligand is a good fit in the narrow 'A' tract of DNA, for two to squeeze in, the groove has to nearly double in width, from 3·2-4Å 'A' tract groove width to 6·8Å, equivalent to the combined drug widths. In the 5′-d(CGCGA^{5}A^{6}T^{7}T^{8}T^{9}CGCG)-3′ sequence (Pelton *et al.*, 1989), the pyrrole rings of distamycin spanned A^5, A^6 and T^7, with the formyl proton pointing in the 5′-direction of the A rich strand (T^8), whilst the second distamycin lies across 5′-A^{6}T^{7}T^{8}T^9-3′ of the 'T' rich strand, indicating that the specificity of sequence and orientation in this 2:1 complex is very high. Blasko *et al.*, (1993) observed a 2:1 and a 4:1 binding mode of distamycin with 5′-d(CGCAAATTTGCG)$_2$, using NMR and molecular modelling. In this case the minor groove opened ~ 16Å to allow these four ligands to bind in their usual antiparallel fashion.

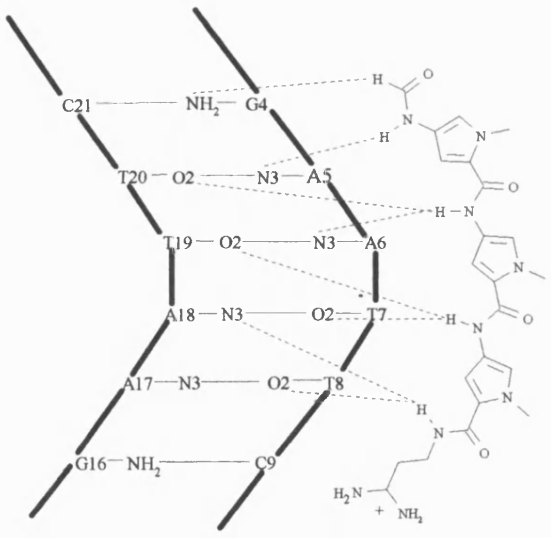


Figure 3 Bifurcating bonds in the distamycin-DNA complex (Pelton *et al.*, 1988).

Garbesi *et al.*, (1996) performed a similar piece of research, with an alternative oligomer 5′-d(CGTACGTACG)$_2$, *via* NMR. The original DNA helical structure was, once

again, preserved on the co-operative binding of two distamycin A molecules. Interestingly, this oligomer does not contain four successive AT base pairs, suggesting that distamycin will interact with mixed sequences. As a consequence of this, the guanine NH_2 moiety may not be responsible for sterically inhibiting the binding of distamycin. It maybe the increase in minor groove width associated with these guanine and cytosine rich GC sequences that yields a 'slack' fit for this ligand, resulting in distamycin's preference for the spatial confinements of the narrow A tracts of DNA. These wider GC grooves are, however, well suited for binding of two distamycin molecules. A large dipole-dipole interaction between the pyrrole and carboxamide groups of adjacent distamycins or netropsins, was detected using X-ray crystallography (Chen *et al.*, 1997). It is this force that is most likely to be responsible for these ligands tolerance for lying next to one another and aiding co-operative binding. This is unlike anything seen in other families of minor groove binding agents.

Following the work of Pelton *et al.*, (1989;1990), Blasko *et al.*, (1993) and Garbesi *et al.*, (1986) which described the anti-parrallel side-by-side binding of two distamycin molecules, Chen and Lown, (1994) manipulated this unique feature to produce cross-linked lexitropsins. Poly(methylene) chains were used to link two tripyrrolocarboxamide strands *via* their central pyrrole nitrogen so as not to interfere with the phasing of binding, or encounter the problem of the formation of a U-turn in the end-to-end bound molecules. Covalent attachment of monomers prevents competition between separate ligands for binding sites. The optimal binding strength K of the dimer is also likely to be greater than the product of stepwise binding (k_a and k_b), suppressing the 1:1 drug: DNA binding mode. CD studies showed that binding to poly(AT)·poly(AT) DNA increased by up to 1000 times in comparison to the monomer, as the length of the methylene linkage increased from four to seven units.

This idea of joining lexitropsin molecules has undergone further development to yield the 'hairpin polyamides' (Fig. 4) (Mrksich *et al.*, 1994) which have been so successful, that they may now be considered as the fifth class of minor groove binders rather than solely an oligopeptide analogue. Modification of the amino acid components allows specific base-pairs to be recognized for example, a selection of eight ring hairpin polyamides were synthesized which differed in the arrangement of pyrrole (Py) and imidazole (Im) amino acids allowing binding to GC bp containing sequences (Swalley *et al.*, 1997). The polyamides ImImPyPy-gamma-ImImPyPy-beta-Dp, ImPyImPy-gamma-ImPyImPy-beta-Dp and ImImImIm-gamma-PyPyPyPy-beta-Dp bound to 5′-TGGCCA, 5′-TCGGCA and 5′-TGGGGA respectively with very high specificity (>250 fold) compared to mismatch sites. The hairpin polyamides were shown to discriminate 5′-GGGG, 5′-GCGC and 5′-GGCC. This recognition of GC base-pairs by hairpin polyamides expands the sequence repertoire by which DNA can be targetted. Parks *et al.*, (1996) also observed lower binding affinities at sites containing single base-pair

mismatches. It was also noted that varying the position of the gamma linker *i.e.* AcImImPy-gamma-PyPyPyPy-beta-Dp and AcPyPyPyPy-gamma-ImImPy-beta-Dp was not found to markedly affect binding affinity and specificity. Surovaya *et al.*, (1997) determined *via* CD studies that the hairpin polyamide bis-netropsin has different structural motifs, depending on whether the oligopeptides are linked tail-to-tail for parrallel side-by-side binding, or head-to-tail for anti-parrallel side-by-side binding. Pt-bis-netropsin was found to traverse 4-5 AT bp of the minor groove in its hairpin form. The chiral properties of these compounds have also been investigated, specifically by chiral substitution of the gamma-aminobutyric acid (gamma) chain, which links the three-ring polyamides consisting of pyrrole and imidazole amino acids, to form a six-ring hairpin (Herman *et al.*, 1998). Footprinting studies demonstrated that the (R)-2,4-diaminobutyric acid linker enhanced the binding affinity by up to 170x relative to the (S) enantiomer, while sequence specificity was affected to a lesser degree increasing only 5-fold at 5′-TGTCA and 5′-ACATT sites. The sequence specificity of the gamma-turn and beta-tail amino acids has also been probed by Swalley *et al.*, (1999), with footprinting studies using the DNA restriction fragments 5′-ATGGCNA and 5′-ANGGCTA. Both gamma-turn and beta-tail amino acids were found to prefer AT/TA base-pairs to GC by 200-400 fold. GC base-pairs were preferred over AT to encase the the 5 bp binding site. This demonstrated that the tail and turn amino acid moieties play an important role in sequence-specificity.

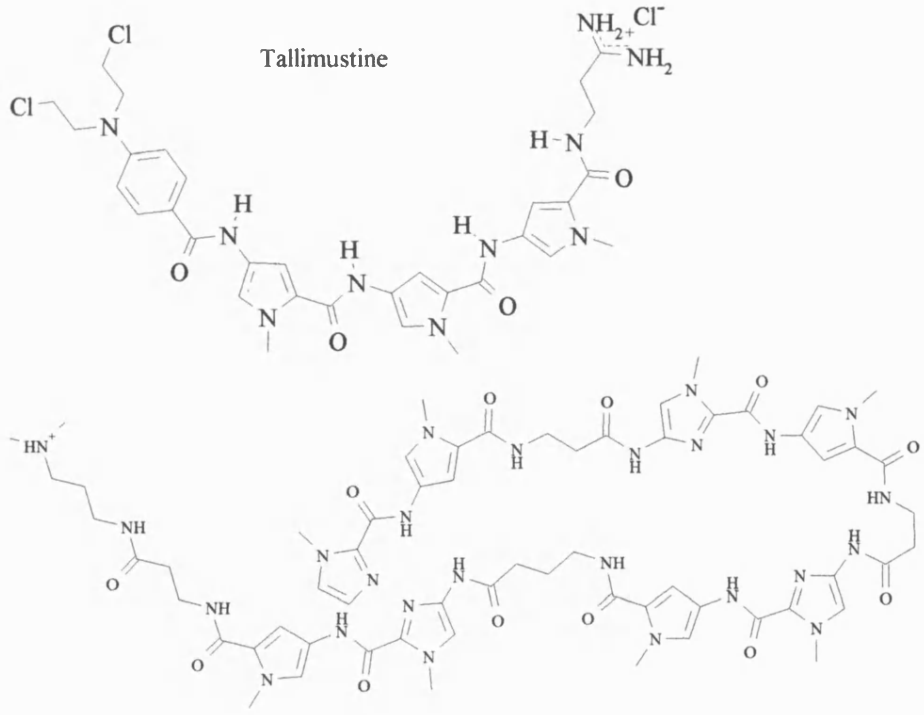


Figure 4 The molecular structures of the hairpin polyamide ImPy-β-ImPy-γ-ImPy-β-ImPy-β-Dp (bottom) and tallimustine (above).

Modification of the individual ring systems using 3-hydroxypyrrole (Hp), N-methylpyrrole (Py) and N-methylimidazole (Im) units in 9- and 11- ring hairpin polyamides allows discrimination between CG (Py/Im), GC (Im/Py), TA (Hp/Py) and AT (Py/Hp) base-pairs (White *et al.*, 1997,1998; Keilkopf *et al.*, 1998). Increasing the length of polyamides so that they recognize 6 or 7 bp sites which is favourable in view of enhanced sequence selectivity, encounters many problems in practise. Spacers between lexitropsins have to be introduced to rephase the hairpin polyamide to achieve sequence recognition, while the subunit 3-hydroxypyrrole achieves selectivity in a negative manner *i.e.* through aversion rather than affinity towards specific sites, which is not an attractive property for drug design (Thurston, 1999). Hairpin polyamides have recently been used to target the HIV-1 transcription element in order to block HIV replication. These ligands prevent host cell transcription factors from binding to the HIV enhancer element (Dickerson *et al.*, 1998). Hairpin polyamides were most effective when a variety of these ligands were combined to competitively inhibit a number of minor groove DNA binding proteins simultaneously.

Tallimustine is an aniline mustard derivative of distamycin (Fig. 4) which is presently in Phase II clinical trials as a broad spectrum antitumour agent. Alkylation is thought to take place in the minor groove at a purine (Pu) N3 in the consensus sequence T^4GPu, where Pu is preferably A (Wyatt *et al.*, 1995) according to modified DNA sequencing techniques. The minor groove comprised of this sequence progressively expands in the 5′- to 3′- direction as it comes out of the narrow propellor-twisted T tract, causing unwinding of the helix. This is an important factor in stablization of the tallimustine adduct, because the widening groove at the guanine base harbours the phenyl ring more comfortably (Ragg *et al.*, 1998). The charged amidinium group lies at the beginning of the 5′-T tract.

Various analogues of distamycin and netropsin have been synthesized (Krowicki *et al.*, 1988; Dwyer *et al.*, 1992; Animati *et al.*, 1995) and their interactions with DNA have been evaluated. Krowicki *et al.*, (1988) replaced the guanidinium group with alkylating moieties such as cyclopropyl, chloro or fluoroacetyl and compared their biological activities with those of unmodified distamycin and netropsin. Cytostatic activity of some of these compounds, notably the chloroacetyl derivative, increased by up to 45 times over netropsin and 18 times that of distamycin. Dwyer *et al.*, (1992) synthesized an imidazole derivative of distamycin which is capable of binding to the consensus sequence 5′-d(AAGTT). The problem of the positive imidazole being adjacent to AN3 was erradicated by substituting the site with a guanine, whose C2 amino protons add a sequence selectivity aspect to the binding, because of their hydrogen-bonding capabilities. Animati *et al.*, (1995) studied the effect of varying the number of pyrrole rings in distamycin. As previously described with distamycin, drug to DNA complexes in a ratio of 2:1 were formed. The tetrapyrrole derivative showed increased specificity at the A_3T_3 binding site, recognizing five base-pairs and exhibiting no sliding from the 5′-AAATT site.

Distamycin and netropsin have great potential as leads in the development of new lexitropsins with enhanced biological activity by logical structural modifications.

1.1.2 The bis-benzimidazoles

Hoechst 33258 was first synthesized by Loewe and Urbanietz (1968). It has subsequently found a diverse number of uses: as a fluorescent histochemical stain, clinically as an anthelmintic agent for treatment of intestinal worms (Raether and Lammler, 1971), and has displayed some activity against intraperitoneally implanted L1210 and P388 leukemias in mice (screening data on Pibenzimol, NCI, Bethesada, USA). Hoechst 33258 consists of three aromatic ring systems; a phenol, two benzimidazole rings and a piperazine ring. These are arranged in a crescent shape (Fig. 5) similar to distamycin and netropsin, perfect for binding into the minor groove of right-handed helical DNA. There are two possible binding modes for Hoechst 33258 to β-form DNA, which are dependent on its concentration (Bontemps *et al.*, 1975). At low ligand concentrations, Hoechst 33258 binds tightly (binding constant ~10^6-$10^7 M^{-1}$) to the outside of the double helix, with its axis at 45° to that of the oligomer, which characterizes association within the minor groove. Higher drug:DNA ratios lead to lower affinity interactions (binding constant 10^4-$10^5 M^{-1}$) with the external surface of DNA. The high affinity binding which occurs in A-T rich regions of the minor groove (Ridler and Jennings, 1980), is comparable to that of the oligopeptide antibiotics. The NH moieties with the hydrogen-bonding capabilities in both families are situated on the concave face, with a cationic group available for favourable electrostatic interactions with the negative potential of the AT groove. The high selectivity of Hoechst 33258 binding has been researched extensively, using a variety of spectroscopic methods and oligomer sequences of the ilk 5′d(CGCXXXXXXGCG)$_2$, where the Xs represent a string of purine base pairs. It is hoped that these will lead to an understanding in the chemistry of DNA recognition by base perturbations in the binding site.

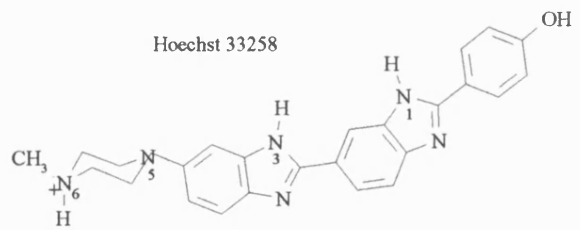


Figure 5 Hoechst 33258 at neutral pH (one positive charge).

Searle and Embrey (1990) investigated the interaction of Hoechst 33258 within the minor groove of 5′-d(CTTTTGCAAAAG)$_2$, by NMR spectroscopy. A single species resulted from the reaction consisting of two Hoechst 33258 molecules binding symmetrically within the

A and T tracts. Both N-methylpiperazine groups were orientated towards the central GC of the complex, illustrating the high level of competence this ligand shows in its ability to 'read' the minor groove. As with distamycin and netropsin, there is evidence of bifurcating bonds between the NHs of the benzimidazole rings and adenine N3 or thymine 02, resulting from characteristic A tract base propellor twisting and a narrower groove. Hoechst 33258 sits within the limits of the AT base pair region, where the electrostatic attraction between them is greatest, contributing to the strength of its binding. Evidence of many van der Waals interactions between the concave face of the drug and DNA was noted. These were thought to be a product of the close complementarity of Hoechst 33258 with the minor groove.

Parkinson *et al.*, (1990) also used spectroscopic methods to look at Hoechst 33258-5′-d(CGCGAATTCGCG)$_2$ interactions. The greatest chemical shift changes belonging to the sugar H1′ protons were used to locate the ligand position, followed by NOE analysis to produce a three-dimensional model. This structure was in agreement with the findings of Searle & Embrey (1990) and an earlier X-ray crystal structure, produced by Teng *et al.*, (1988). Other studies have also targetted the 5′-AATT with varying flanking sequences. Fede *et al.*, (1991) decribed the positioning of the Hoechst 33258 subunits in detail within the 5′-d(GTGGAATTCCAC)$_2$ duplex. The piperazine moiety in this unsymmetrical complex was located at A^5T^6 and $T^{19}T^{20}$ by intermolecular NOE data generated between this group and the AH2 and sugar H1′ protons. Hydrogen-bond donors and acceptors lock the ligand into juxtaposition with the binding site, contributing to the stability of the complex. The NH of benzimidazole adjacent to the phenol head is bonded with T^7O2, while the neighbouring benzimidazole interacts with T^8O7. Embrey *et al.*, (1993) used a sophisticated DNA duplex 5′-d(GCTAATTACC)$_2$, consisting of five consecutive binding sites, containing the minimum 3xAT base pairs, to react with Hoechst 33258. Once again binding only occurred at the 5′-AATT site, phenyl ring protons being situated closely to A^4H2. The NOE data suggest that the final tightly bound structure is analogous to those described previously. Hoechst 33258 displays no partiality for GC flanking sequences which could have possibly accommodated the bulky piperazine ring within the wider groove with greater ease. Intramolecular NOEs indicated no perturbation in β-helical DNA structure, so the flanking 5′-AT bases are able to harbour a piperazine bby some subtle change in minor groove width.

The analogue meta-hydroxy Hoechst 33258 was allowed to react with 5′-d($C^1G^2C^3G^4A^5A^6T^7T^8C^9G^{10}C^{11}G^{12}$)$_2$ to assess the effect of the hydroxyl group on selectivity and binding, by investigating any new hydrogen bond-formations (Parkinson *et al.*, 1994). In agreement with earlier results, the piperazine ring was situated at T^8 of the 5′-d(AATT^8) binding site (from H4′ NOE data). Embrey *et al.*, (1993) noted that the phenol ring flips in the minor groove even at low temperatures, although this activity ceases on relocation of the hydroxyl

group to the meta position. This led Parkinson *et al.*, (1994) to predict from their model that hydrogen-bonds between the meta-hydroxyl and C=O of C^9 or the exocyclic NH_2 of G^4 were in place and were substantial enough to prevent ring movement. Additional hydrogen-bonds would undoubtably increase the stability of the complex, plus enhancing toleration and possibly recognition of a GC base pair at the 5′- end of the AATT target sequence. However, as previously discussed, hydrogen-bonds are not directly responsible for the specificity with which these ligands can target a sequence. Both pyrimidine and purine bases having equal numbers of hydrogen-bond acceptor atoms; O2 and N3 respectively.

An X-ray crystal structure of the meta-hydroxy analogue of Hoechst 33258, was determined with the same 5′-d(CGCGAATTCGCG)$_2$ sequence (Clark *et al.*, 1996). This compound has its benzimidazole group replaced by pyridoimidazole. This meta-hydroxy analogue was found to span 5′-AATTC. The intermolecular hydrogen-bonding influenced the conformation of the whole structure. The phenol adjacent to A^5 within the minor groove, was present as 'in' and 'out' conformations relative to the minor groove. Hydrogen-bonds were formed in each conformation, although the latter case required extended water networks to achieve this. The customary bifurcating H-bonds along the concave surface of the ligand were present, but the pyridyl nitrogen N7 did not take part in this, so the overall result was very similar to that of binding of Hoechst 33258 itself. For these subunits to align for H-bonding simultaneously within the minor groove, the ring systems have to twist relative to each other away from co-planarity. This results in widening of the groove and, hence, a different overall structure to that of Hoechst 33258.

Wood *et al.*, (1995) examined the crystal structure of another analogue of Hoechst 33258, where an imidazole ring replaces piperazine. The main difference between this model and that of Clark *et al.*, (1996) is the narrowing of the minor groove at the 3′-end of the binding site, relative to the substitution with the less spatially consuming imidazole group. This is thought to be the main reason for the increase in the binding affinity of this analogue compared to Hoechst 33258, because both molecules form the same hydrogen-bonds in the groove. Clark *et al.*, (1997) replaced the piperazine and phenyl groups of Hoechst 33258 with an amidinium and phenylamidinium moiety, respectively. This demonstrated that, where no other interactions were available, five hydrogen-bonds were sufficient for high affinity binding.

Clearly, narrowing of the minor groove width (associated with a close complementary fit around the drug), and hydrogen-bonding, are both necessary for high affinity binding of Hoechst 33258. In all stuctures, regardless of the sequence, there is one bifurcating hydrogen-bond from one benzimidazole NH, with either a single or bifurcating H-bond from the second benzimidazole ring. Abu-Daya *et al.*, (1995) conducted a study using DNase I footprinting

techniques to probe the binding of Hoechst 33258 to oligomers of varying A·T target sequences. These results suggested that future novel compounds should be based on the benzimidazole structure, due to its significantly higher affinity for AATT sites than other arrangements of A and T by berenil or the oligopeptide ligands.

1.1.3 Bisamidines

Berenil is a member of a large group of compounds, called the bisamidines, which have a common terminal amidine moiety. First synthesized in the 1950s, it was found to possess a variety of useful activities including antimicrobial, antiviral and antiparasitic properties (Newton, 1975). It is still used as a trypanocide in Africa today as a treatment for bovine trypanosomiasis, although its effectiveness has decreased over the years due to resistance (Kopka & Larsen, 1992). Berenil is also capable of inhibition of mammalian topoisomerase II activity and binding to the DNA-containing kinetoplasts in mitrochondria (Portugal, 1994). The molecular structure of berenil (Fig. 6), consists of two diamidine benzyl rings joined by a triazene bridge. Braithwaite and Baguley (1980) demonstrated that berenil bound to the minor groove of DNA duplex, selectively at dA·dT base pairs. This was also proved *via* displacement of ethidium bromide from its intercalation site in DNA, by a 5-fold preference over dG·dC sequences (Baguley, 1982).

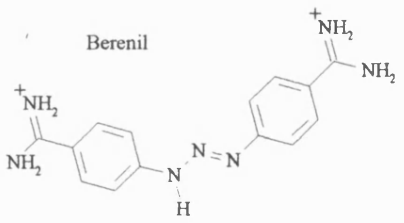


Figure 6 The molecular structure of berenil.

As for many other DNA-drug complexes, the binding of berenil to a selection of $p(dA)_n \cdot p(dT)_n$, $p(dAT)_n \cdot p(dAT)_n$ and $p(dTA)_n \cdot p(dTA)_n$ DNA duplexes, has been investigated by X-ray crystallography (Pearl *et al.*, 1987). Analogously to the oligopeptides and bis-benzimidoles discussed previously, berenil binds isohelically within the A-T minor groove. The concave surface of berenil is complementary to the base of the minor groove. The presence of a guanine base, with an amino moiety protruding into the binding region, would sterically hinder berenil from binding in a G-C rich locality. Both positively charged terminal amidine groups of berenil are able to rotate. This enables them to come into close hydrogen-bonding contact with the oxygen atoms of the phosphate backbone, although this is at the expense of the hydrogen-bonding within the base pair. The ring systems in the molecular model are slightly twisted relative to one another by approximately 9·6°. Two strong hydrogen-bonds were formed in the

$d(AT)_4 \cdot d(AT)_4$–berenil complex, between the ligands terminal amidine groups and O2 of thymine. In constrast, the reverse occurred with the $d(TA)_4 \cdot d(TA)_4$ duplex, the hydrogen-bonds being directed towards the N3 of adenine acceptors instead. Adenine H2 obstructs this latter interaction, providing a possible explanation as to why berenil has a higher affinity for 5′-AT rather than 5′-TA sequences.

Yoshida *et al.*, (1990) generated a molecular model from NOE data collected on the berenil–5′-d$(GCAATTGC)_2$ complex, which conflicts with the favoured binding model proposed by Pearl *et al.*, (1987). This solution structure agreed with the proposition that berenil bound spanning the central 5′-AT base-pair. However, Yoshida *et al.*, (1990) placed the ligand effectively in a mirror image orientation. This allowed hydrogen-bonding to occur between the internal adenine N3s and the amidine protons. Unfortunately, this model does not take into account the length of the berenil molecule and the fact that it is too long to span comfortably the width of the minor groove across an AT base pair without undergoing some structural deformation (Kopka and Larsen, 1992). Brown *et al.*, (1990) examined the crystal structure of 5′-d$(C^1G^2C^3G^4A^5A^6T^7T^8C^9G^{10}C^{11}G^{12})_2$–berenil complex. The binding sequence was found to consist of a 5′-AAT section, where the amidinium tail was hydrogen-bonded to A^5N3 and A^6O4′ *via* a water molecule, whilst the head group formed a bifurcating bond to A^{18}N3 and T^8O2. This model accomodates the size of berenil correctly. Lane *et al.*, (1991) studied the same complex by NMR and established that the binding site was 5′-ATT, which is an equivalent site on the opposite strand to that previously described (Brown *et al.*, 1990). Molecular mechanics calculations produced a structure in which the berenil was planar and isohelical with the minor groove floor, allowing for hydrophobic interactions between phenyl rings and the walls of the groove. The charged amidinium groups formed hydrogen-bonds with the N3 of A^6 and A^{17}, when the triazene moiety aligned itself NH-N=N across 5′-$A^6T^7T^8$, to minimize deformation of the DNA helix.

Hu *et al.*, (1992) have studied berenil binding by NMR, using a 5′-TA-containing sequence 5′-d$(GCTTAAGC)_2$ According to Pearl *et al.*, (1987) and Abu-Daya *et al.*, (1995), this 5′-TA step was less favourable in comparison with the 5′-AT step for ligand binding. The greatest chemical shift was observed for thymidine protons, whilst, in contrast, it was the adenosine protons in the 5′-AATT complex that experienced a large chemical shift. This suggested that berenil was in the same orientation in both sequences. From the NOE data the binding site of 5′-d(GCTTAAGC) was deduced as 5′-TAA. Smaller chemical shift differences between this complex and free duplex, compared to those seen with the 5′-AATT sequence were indicative of weaker interactions. Hence, this sequence displays a lower affinity for berenil than 5′-d$(GCAATTGC)_2$, in agreement with both Portugal and Waring (1987), and Pearl

et al., (1987). Interestingly, berenil binding to 5′-d(CGCAAATTTGCG)$_2$ also occurs *via* hydrogen-bonds to the O2 of thymine, around the central four base-pairs (Brown *et al.*, 1992).

More recent research by Pilch *et al.*, (1995) has probed the ability of berenil to intercalate with DNA as it simultaneously binds to the minor groove. Viscometry of the berenil-poly(A-T)$_2$ complex indicated that some intercalation was taking place, because an increase in relative viscosity is proportional to an increase in molecular length, a consequence of intercalation. Unwinding of the negative supercoils of the pBR322 plasmid was observed, which is only achievable by insertion of berenil between the bases of the helix. Two inflection points were visible on the CD spectra which correlated with more than one optically detectable ligand binding event. It was also observed in UV melting studies that berenil binding to p(A-T)$_n$ was less exothermic and hence less favourable than netropsin (minor groove binder) or ethidium bromide (intercalator), suggesting that the binding mechanism could be a mixture of both modes. Pilch *et al.*, (1995) suggested that berenil does not bind to DNA solely *via* the minor groove binding mode, as all previous paths of research have indicated. It maybe that there is a mixed binding mode; intercalation versus minor groove binding in competition with one another, or possibly a single binding mode with dual characteristics. Correlations between mixed binding modes and anticancer activity by topoisomerase I inhibition (Chen and Liu, 1994), mean that berenil could be a lead for future drug design if it is capable of both intercalation and minor groove binding.

DAPI (4′,6-diamidino-2-phenylindole) is another diarylamidine compound synthesized in the 1970s, as the result of a comprehensive search for more effective trypanocides. This novel unfused aromatic compound now has the reputation of being a classic minor groove binder. Like berenil, DAPI has head and tail diamidino moieties which are available for hydrogen-bonding. The centre of the molecule consists of a benzene ring connected to an indole ring system (Fig. 7). DAPI is an inhibitor of a selection of DNA processing enzymes such as RNA polymerase II, DNA ligase, exonuclease III and DNA polymerase I (Parolin *et al.*, 1990). Chiang *et al.*, (1994) found DAPI to be responsible for inhibition of the TATA binding protein TBP, from binding to the TATA box consensus sequence, which is a prerequisite for RNA polymerase II binding and hence gene transcription. DAPI is fluorescent, which has led to its use clinically as a chromosomal stain in screening cancer cells for chromosomal abnormalities.

Larsen *et al.*, (1989) analysed the X-ray crystal structure of DAPI bound to 5′-d($C^1G^2C^3G^4A^5A^6T^7T^8C^9G^{10}C^{11}G^{12}$)·5′-d($C^{13}G^{14}C^{15}G^{16}A^{17}A^{18}T^{19}T^{20}C^{21}G^{22}C^{23}G^{24}$) This ligand was found to bind in a single orientation across the central AATT segment. A bifurcating hydrogen-bond from the N3 of amidinium adjacent to the indole was formed with T^{20}O2 and

A^6N3, whilst N4 of the amidinium tail hydrogen-bonded with the sugar oxygen of A^{18}. The nitrogen of the indole ring forms a bifurcated hydrogen-bond with $T^{19}O2$ and T^7O2. Therefore, DAPI adheres tightly to the 5′-ATT binding site within the minor groove.

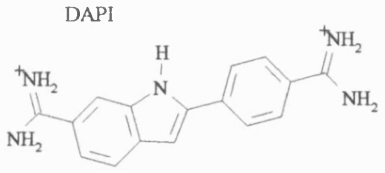


Figure 7 Molecular structure of DAPI.

It has also been demonstrated that DAPI is, in fact, a mixed mode binder. Wilson *et al.*, (1989;1990) provided evidence of intercalative binding within d(G·C) and mixed d(G·A·C·T) sequences. Changes in base sequence composition are responsible for geometric changes to the minor groove, namely the width and depth, by becoming wider and shallower. As a consequence of this, the ligand binds to the minor groove with low affinity and the intercalative binding mode becomes more energetically favourable.

Trotta *et al.*, (1996) discovered that the $d(AT)_n$ base pairing was not the only factor to determine whether DAPI bound to the minor groove of DNA. The ligand:nucleotide binding ratio also has an important effect. NMR data on the DAPI-5′-$d(CGATCG)_2$ complex, which existed at a high 2:1 drug:duplex binding ratio, provided evidence for two simultaneous modes of binding. Firstly, there is binding within the minor groove across 5′-ATC, and secondly, there is π-π stacking of the aromatic ring systems of DAPI onto the terminal base pairs. This mode of external stacking is novel but not so useful biochemically. This is because genes are surrounded by nonsense codons, which are DNA sequences void of information for protein transcription. It is at the end of one of these regions where DAPI would stack.

Propamidine and pentamidine are probably the most comprehensively researched aromatic diamidines after berenil. X-ray crystallography and NMR spectoscopy have been used to try to find the key, to these ligands binding modes and sequence specificities. A large databank has been compiled on their complexes with various oligonucleotide sequences. The structure of pentamidine or 1,5-bis(4′-amidinophenoxy)pentane is given in (Fig. 8). Propamidine or 1,3-bis(4′-amidinophenoxy)propane (Fig. 8) is the short chain homologue of pentamidine. Clinical interest in these bisamidines has revived due to pentamidines application in the treatment of *Pneumocystic carinii* pneumonia, an opportunistic infection occurring in greater than 70% of patients with AIDs, which is a major cause of their death (Tanious *et al., 1997)*. Design of new ligand analogues is important because of toxicity. Like the other members in this family, pentamidine and propamidine are also antitrypanosomal and antiviral agents.

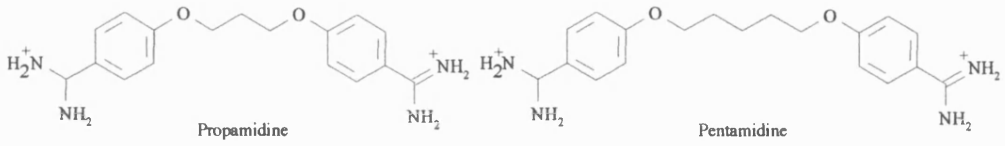


Figure 8 Molecular structures of propamidine and pentamidine.

Jenkins *et al.*, (1993) and Greenidge *et al.*, (1993) used NMR and molecular modelling studies to probe the binding of pentamidine to 5′-d(CGCAAATTTGCG)$_2$ and 5′-d(CGCGAATTCGCG)$_2$, respectively. Three possible binding sites were available in the 5′-AAATTT sequence, taking into account pentamidines requirement for a 4-5 base-pair segment for binding (Zimmer and Wahnert, 1986). Preferential binding took place across the asymmetric 5′-ATTT site, where van der Waals forces and hydrogen-bonds from pentamidine terminal amidine NH_2 to A^6 and $A^{16}N3$, are more abundant. The molecule retains its planarity and is a good isohelical fit within the minor groove, so that binding energy penalties are not incurred by DNA distortion. Molecular models generated by Greenidge *et al.*, (1993) agree with these conclusions; in that pentamidine is more elongated than in the crystal structure, spanning over four bases. Analysis of some pentamidine analogues demonstrated that varying the number of (methylene units)$_n$, where 'n' is an odd number in para-substituted analogues or even in metasubstituted ones, increased the ability of these molecules to fit isohelically within the minor groove.

Conte *et al.*, (1995) used the asymmetric sequence 5′-d(GCAATGAGCG) to bind propamidine. 5′-AAT was found to be the major binding site although some binding also occurred across 5′-TGA indicating that a GC base pair can be tolerated in an AT rich target sequence. Propamidines binding site is 3-4bp in length, 1bp shorter than the 4-5bp which pentamidine spans, reflecting propamidines shorter linker. Nunn and Neidle, (1995) compared the crystal structure of propamidine binding to 5′-d(CGCGAATTCGCG)$_2$ with 5′-d(CGCAAATTTGCG)$_2$. The binding differed by a 2Å displacement of the ligand toward the 3′-end of the A_3T_3 strand, relative to the position in the A_2T_2 sequence. This is enough to redefine hydrogen bonding patterns as a consequence of the modified minor groove geometry *i.e.* increased length of the AT sequence ($A_2T_2 \rightarrow A_3T_3$), leads to a narrower groove. This results in more van der Waals interactions and hence, a more 'comfortable' fit for propamidine in the groove.

Jenkins and Lane, (1997) stated that the binding affinity of berenil>propamidine>pentamidine. This is probably a reflection of the entropic cost endured on binding, in a bid to procure the 'best fit' conformer for the minor groove. This suggests future

rational design should move in the direction of ligands possessing less internal rotation for increased binding affinity.

1.1.4 Bis(quaternary ammonium)heterocycles

The bis(quaternary ammonium) heterocycles (BQAH) are a family consisting of a synthetic series of molecules which bind to the minor groove of DNA. They are related to polyamidines, but have quaternary ammonium heterocycles as the charged centres. Perhaps the most well known analogue is SN-6999 *i.e.* 4-[4-[4-(4-quinolylamino)benzamido]anilino]pyridine, which consists of four aromatic rings: a quinolinium (Q), benzamido (BQ), anilino (BP) and a pyridinium (P) (Fig.9). Other members are SN-4094, SN-6136 and SN-18071 (Fig.9). BQAHs exhibit potent antitumour and antiviral activies (Cain *et al.*, 1969) *in vivo* and *in vitro*, mainly as a result of their effects on DNA metabolism and inhibition of DNA gyrase (Storl *et al.*, 1993). The lipophilicity of these drugs is thought to have a positive correlation with anticancer potency. While $p(dAdT)_n \cdot p(dAdT)_n$ rich DNA sequences, are preferred over $p(dGdC)_n \cdot p(dGdC)_n$ (Denny *et al.* 1979; Baguley 1982). Some of the early analogues, *e.g.* SN-4094, showed chronic toxicity *in vivo*. There are little mutagenicity data available on BQAHs as genotoxins; however, due to its ability to bind to DNA, it is unlikely to be benign (Turner and Denny, 1996).

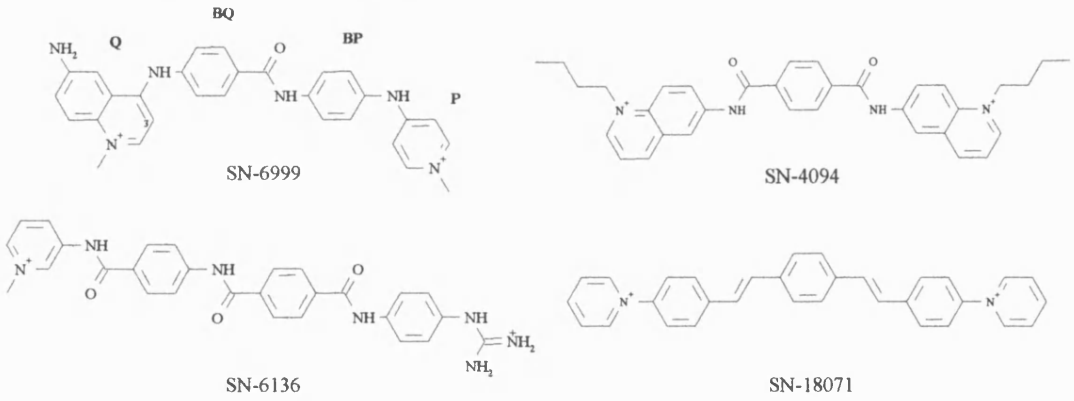


Figure 9 Molecular structures of some bis(quaternary ammonium)heterocycles.

It was established by Braithwaite and Baguley (1980), through viscosity measurements that SN-6999 bound to the exterior of DNA in the minor groove. This agreed with data showing SN-6999 was unable to unwind closed superhelical DNA, the classic test for intercalation. Leupin *et al.*, (1986) investigated the binding of SN-6999 to the DNA duplex 5′-$d(GCATTAATGC)_2$ by NMR. From analysis of the NOEs, it was concluded that the ligand bound in the AT rich region with little pertubation of the β-DNA helix, as all proton walks remained intact. SN-6999 bound to the minor groove in an extended manner, spanning approximately five base pairs. However, because the lifetime of the complex was quite short,

the asymmetric ligand binding did not result in loss of magnetic equivalence between the complementary strands as expected. Temperature-dependent changes were noted for the lineshape of a thymine methyl. Alongside NOE data, this led to the conclusion that SN-6999 bound to two equivalent sites in the duplex (Fig. 10) (Searle 1993).

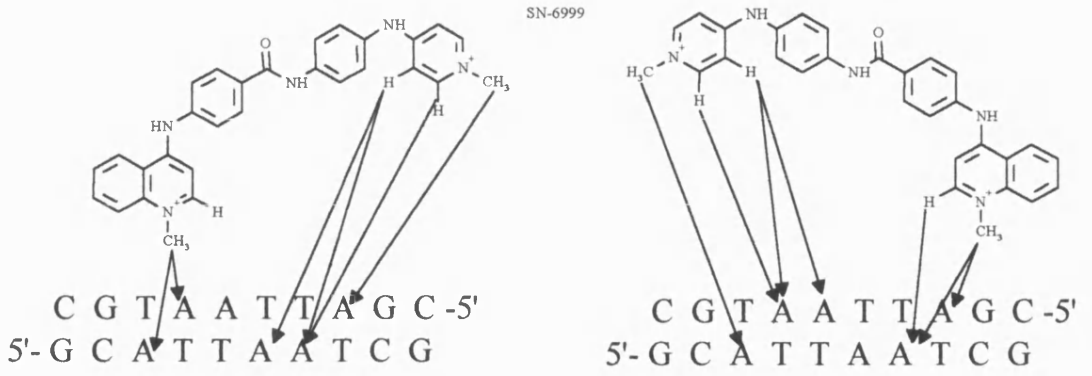


Figure 10 Diagram to show how SN-6999 binds to equivalent sites in the DNA duplex.

Figure 10 shows that these two species are not distinguishable and, hence, the NOEs for each strand are degenerate. Leupin *et al.*, (1986) proposed that this binding mode was a result of a 'flip-flop' process of intermolecular rearrangement, without dissociation of the drug from the complex. This was because only a slow exchange of imino protons was observed, suggesting that SN-6999 was obstructing solvent from accessing this area of base-pairing in the DNA minor groove.

Gao *et al.*, (1993) presented a crystal structure of SN-6999-5′-d($C^1G^2C^3[e^6G^4]A^5A^6T^7T^8C^9G^{10}C^{11}G^{12}$), in which SN-6999 did not exhibit such a pronounced quarter-moon shape, as compared to other minor groove binders. The P and Q rings were not in plane with the BQ and BP rings, which were orientated 'edge on' within the minor groove 5′-AA sequence. This provided the opportunity for abundant van der Waals interactions. The P ring was situated at $e^6G^4{\cdot}C^{23}$ and $C^3{\cdot}G^{22}$, whilst Q was adjacent to the centre of the sequence. The positive charges of both moieties contributed to the ligands affinity for the DNA. SN-6999 binding also results in a tilted and stiffened DNA helix. By analogy with other minor groove binders, initial specificity was attributed to H-bonding ability (Baguley, 1982). However, SN-18071 defies this proposal because it is unable to form any hydrogen-bonds, yet has a binding affinity only slightly lower than SN-6999.

Chen *et al.*, (1992) extended the work of Leupin *et al.*, (1986) to a piece of DNA with a shorter AT sequence in an attempt to produce a 1:1 complex with SN-6999 bound in only one orientation. Guanine amino groups were placed on each side of the target sequence (5′-d(GGTTAATGCGGT)·5′-d(ACCGCATTAACC)) to impede sterically any ligand sliding. The

NMR chemical shifts were greatest for bases $T^3 \rightarrow G^8$ on one strand and from $A_{18} \rightarrow C_{23}$ on the other opposite strand, indicating the position of the bound SN-6999. Although no multiple sets of signals were visible, the methyl resonances were broad, implying that two orientations of the drug were present, exchanging at an intermediate rate. The proportionate NOE intensities of the quinolinum $3H \rightarrow A^{18}$ and C^{23}, at opposite ends of the binding site, demonstrated that these two drug-DNA species existed in equal quantities.

Rydzewski *et al.*, (1996) hypothesized that a homogenous 'A' tract binding site 5′-d($G^1G^2G^3A^4A^5A^6A^7A^8C^9G^{10}G^{11}$) could hold the key to trapping SN-6999 in one orientation, due to the imposed spatial confinements. In fact, the binding ratio was only 7:3 in favour of the pyridinium ring being situated near 5′-A^5. This preferential binding mode optimized shape complementarity and the various types of interactions between DNA and SN-6999, resulting in this observed 7:3 major:minor orientational ratio. SN-6999 associated more closely with the 3′-ends of both the $p(A)_5$ and the $p(T)_5$ sequences, in effect lying obliquely between both strands as would a covalent cross-linker. A high resolution structure of a BQAH-DNA complex is still required to elucidate the molecular basis, kinetics and affinities of the binding of these ligands before rational design of new, related therapeutic compounds is really possible.

1.2 The covalent minor groove binders

There are two major classes of covalent minor groove binders: the pyrrolo[2,1-*c*][1, 4]benzodiazepines and the cyclopropapyrroloindoles. Their modes of binding and specificity will now be reviewed in detail.

1.2.1 Pyrrolo[2,1-c][1,4]benzodiazepines

The pyrrolo[2,1-*c*][1,4]benzodiazepine (PBD) family is composed of natural products from various thermophilic *Actinomycetes.* They are sometimes known as 'antitumour antibiotics' owing to their capacity to inhibit strongly the multiplication of *e.g.* mice reticulo-endothelial sarcoma cells at low drug (*e.g.* sibiromycin) concentrations. At high drug concentrations, the growth of various bacteria, such as *Bacillus mycoides,* is also inhibited (Gause *et al.*, 1969). Anthramycin was isolated from *Streptomyces refuineus* (Tendler and Korman, 1963) and the name reflects the anthranilic acid component of the molecule. Sibiromycin was discovered in 1969 in Siberia (Gause *et al.*, 1969), from *Streptosporangium sibiricum.* Tomaymycin was derived from *Streptomyces tomaymyceticus,* in the Tomei region between Tokyo and Nagoya (Arima *et al.*, 1972). Lastly, the neothramycins A and B were

isolated from *Streptomyces N° MC916-C4* (Takeuchi *et al.*, 1976), also in Japan. All these compounds have a common structural feature the pyrrolo[2,1-*c*][1,4]benzodiazepine nucleus. However, they differ in the degree of pyrrole ring saturation, the side chain at C2, the pattern of substitution and the substituents on the aromatic ring (Fig. 11). For example, anthramycin and sibiromycin are both 8-methoxy-9-hydroxy substituted, whilst tomaymycin is 7-methoxy-8-hydroxy. Sibiromycin has a large sugar '7-sibirosaminide' side-chain. The carbinolamine at N10-C11 is the functional group capable of reaction with DNA. In the case of the neothramycins where a 10,11 enamine is present, hydration in solution results in the required carbinolamine.

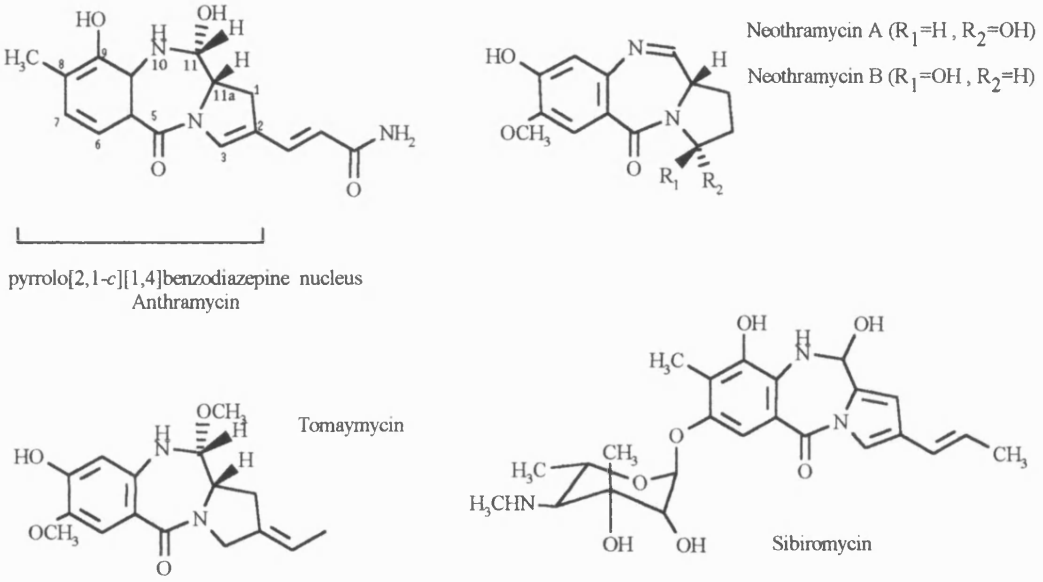


Figure 11 Molecular structures of the pyrrolo[2,1-*c*][1,4]benzodiazepines

All pyrrolo[2,1-*c*][1,4]benzodiazepines are potent inhibitors of RNA and DNA synthesis, whereas the synthesis of proteins is not affected (Horwitz and Grollman, 1968; Maruuama *et al.*, 1978; Gause and Dudnik, 1971). On anthramycin binding to DNA, some cells try to remove the ligand by an excision process, resulting in induction of unscheduled DNA synthesis (Hurley *et al.*, 1979). Human skin fibroblasts use this process and can remove approximately 86% of bound anthramycin in 72 hours. It is this inhibition of nucleic acid synthesis which is responsible for the very potent antitumour activity of the PBDs, as demonstrated *in vivo* (Hurley, 1977). However, their clinical application has been prevented by dose limiting toxicities, such as cardiotoxicity in the case of anthramycin and sibiromycin (Cargill *et al.*, 1974), and tissue necrosis at the site of injection. The ability of this family of compounds to cause mutation has also been investigated using strains of *Salmonella typhimurium* and *Saccharomyces cerevisiae* (Hannan *et al.*, 1978). Anthramycin does not alter the mutation rate of these cells, but excision and recombinational repair processes were

necessary for elimination of PBD adducts from their genetic material. For a review, see Turner and Denny, (1996).

Hurley *et al.*, (1977) utilized radiolabelled pyrrolo[2,1-*c*][1,4]benzodiazepines, in order to probe the structure of the resulting adduct and the mechanism of action which leads to it. Anthramycin, sibiromycin and tomaymycin were used to elucidate how their structural differences affected their reactivity. Each of these compounds was incubated with calf thymus DNA. Sibiromycin reached saturation within 20 minutes, whereas in comparison, tomaymycin and anthramycin were slower to bind, taking three times as long. Competition experiments were performed to ascertain whether all compounds bound to the same DNA sequence. In conclusion, it was found that sibiromycin was in complete competition with the binding of tomaymycin and anthramycin. Incomplete competition between anthramycin and tomaymycin, meant that anthramycin-saturated DNA was still able to bind 30% of the tomaymycin required to saturate DNA alone. Therefore, at least three possible binding sites exist;

1) selective for anthramycin and sibiromycin
2) selective for tomaymycin and sibiromycin
3) anthramycin, tomaymycin or sibiromycin can bind

Adjusting the pH to an acidic value resulted in destabilization of the PBD-DNA complexes. Sibiromycin was able to withstand the most acid conditons. At the point where the complex became unstable, dissociation occurred and the ligand was released in its original state. At physiological pH, this could effectively lead to drug walking along the DNA, due to repeated cycles of binding, excision and dissociation. The hierarchy for acid stability, sibiromycin>anthramycin>tomaymycin, correlates with the rate of reaction. This suggests that sibiromycin binds very tightly to the DNA preventing denaturation *via* acid hydrolysis. Radiolabelling of non-exchangeable hydrogens in the pyrrole ring and side chains established that none of these were involved in the mechanism of binding these ligands to DNA. This led Hurley *et al.*, (1977) to propose a mechanism of reaction where the electrophilic C11 of the N10-C11 imine undergoes nucleophilic attack by the N2 of a guanine base to form an aminal (Fig. 12).

Corey, Pauling and Koltun (CPK) space-filling models of sibiromycin, tomaymycin and neothramycins A and B DNA adducts were constructed by Hurley *et al.*, (1980). As with the minor groove binders previously discussed, the concave surface of the PBD, carrying protons 9,10,11,11a and 1, fit in a manner complementary to the base of the minor groove. The models demonstrated that these drugs positioned themselves in the groove, without giving rise to helical pertubations. They probably stablize themselves through hydrogen-bonding from the phenolic protons to the 2-keto group of cytosine or thymine. This could be to the same base-pair to which anthramycin is bound, or in the case of sibiromycin to the adjacent base-pair,

which could also be an adenine or guanine N-1 of the modified strand. It is this adjacent base pair 2-keto group which can also accept hydrogen-bonds from the N10 proton of anthramycin and tomaymycin. The N10-H of sibiromycin H-bonds to the cytosine of the modified base pair because of the altered 7-membered ring conformation, resulting from C11a being sp^2 instead of sp^3 hybridized. Sibiromycin is the only PBD not to have an asymmetric carbon at C11a, which is responsible for an advantageous right-hand twist of the molecule. This allows the PBD to form a perfect complementary fit within the minor groove of β-helical DNA (Mostad *et al.*, 1978). It is this 'S' configuration which is the naturally occurring form of the PBDs. However, the sibirosamine sugar does have a stabilizing influence *via* its interaction with the DNA backbone, increasing the strength of attachment as it protrudes out of the minor groove. Interestingly, adriamycin and daunomycin require an amino sugar moiety for their biological activity, perhaps expressing the significance of the non-covalent interactions of this group with DNA.

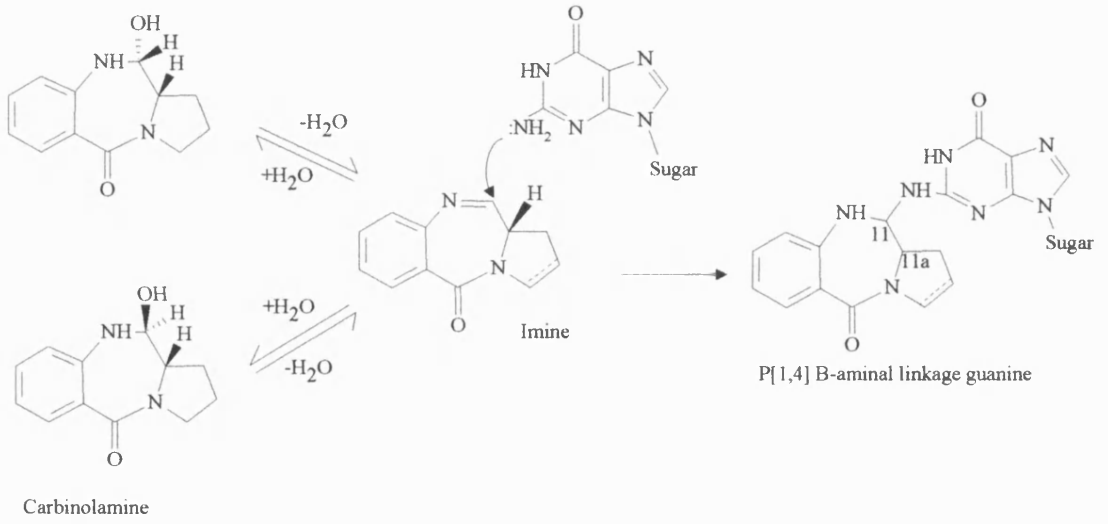


Figure 12 Mechanism of reaction between PBD nucleus and guanine NH_2 of DNA, to form a covalent adduct.

Petrusek *et al.*, (1981) also used CPK models to investigate the molecular structure of PBD-DNA adducts. Saturation binding of these ligands produced a rigid helix, with the drug molecules lying end-to-end within the minor groove, like a third DNA strand. Binding of anthramycin was also revealed as a two-phase process, composed of a fast (20-50min) and a slow (~24h) reaction. The fast reaction occurred with p(dG) sequences, 5′-GG being a preferred binding site. Anthramycin exhaustively bound to these sites prior to resorting to 5′-GX sequences, where X was A, C or T. These small ligands span two to three base-pairs. The stereochemistry at C11 was determined to be 'S'. Neothramycin A and B analogues were used to investigate the steric limitations imposed by the substituents of these drugs on binding to the minor groove. Neothramycin A was found to bind with higher affinity than the B isomer, and

the B isomer C-3-methoxy or C-3-butoxy derivatives. However, these analogues of neothramycin A did not bind. This indicates that C-2 or C-3 substituents facing inside the minor groove, as in the A isomer, can only be tolerated if they are small, unlike in the B isomer where larger groups pointing outwards from the groove can be accommodated. Limiting the size of substituents and consequently of the resulting ligand, is likely to be advantaegous in the design of PBDs. This is because the more they are concealed within the minor groove, the less detectable they are to excision repair complexes.

Further molecular modelling studies used AMBER to evaluate the directional binding preferences, in the minor groove for non-sugar containing PBDs (Remers *et al.*, 1986). The ligands were covalently bound to G^3N2 of 5′-d used AMBER to evaluate directional binding preferences within the minor groove, for non-sugar $(A^1T^2G^3C^4A^5T^6)_2$, and proved to be an excellent fit when positioned in either orientation. Energy analysis suggests that the favoured conformation of the adduct, has the side-chain from C-2 extending alongside the modified strand in the 5′-direction.

The pyrrolo[2,1-*c*][1,4]benzodiazepines are selective for guanine-rich DNA sequences, unlike the AT selective agents previously discussed which prefer the narrow confinements of p(dA) tract DNA. Hertzberg *et al.*, (1986) hypothesized that the PBDs were sensitive to sequence-dependent variations in helical structure, resulting in their sequence specificity. The observation of Hurley *et al.*, (1977) of incomplete competition between anthramycin and tomaymycin, the perceived fast reaction with p(dG) sequences (Petrusek *et al.*, 1981) and the increased inhibition of restriction enzymes on ligand binding to p(dG) (Kaplan, 1982) all testify to this conviction. Hertzberg *et al.*, (1986) used MPE – Fe(II) footprinting techniques to identify the exact PBD binding locations. Cleavage of the DNA took place 2 - 3 bases either side of the guanine, where the ligand was bound and spanned a total of 3 bases. Enhanced cleavage was noted in some sequences where PBD binding had resulted in conformational changes in the nature of the DNA helix. The preferred binding sequence was 5′-PuGPu and the most disliked was 5′-PyGPy. Each PBD has a different twist angle across its nucleus, measured from the aromatic ring to the 5-membered ring *e.g.* anthramycin is 35° and tomaymycin is 9°. They also differ in their degrees of pyrrole ring pucker (Arora, 1981). Anthramycin (larger dihedral angle) binding, was proposed to open up regions of DNA close to the binding sequence, possibly resulting in increased enzymatic cleavage of DNA.

Hurley *et al.*, (1988b) synthesized analogues of the PBDs to investigate the effect of the substituents at C-8 on drug reactivity. These were evaluated using exonuclease III stop assays. Electron-withdrawing substituents, such as chlorine, reduced the availability of the electrons in the lone pair of N10 and protonation of N10 did not take place, rendering the imine

stable to nucleophilic attack. Electron-donating substituents, *e.g.* -OH, $-CH_3$, $-OCH_3$, had the reverse effect, increasing the availability of the lone pair at N10 and consequently the rate of nucleophilic attack on the protonated imine. The extent of DNA alkylation by these drugs correlated with their *in vivo* potency.

The anthramycin–5′-d(ATGCAT)$_2$ adduct has been analysed using ^{13}C and ^{1}H NMR spectroscopy by Graves *et al.*, (1984). On ligand binding to this palindromic sequence, the DNA strands lose their equivalence, as expected. The resonance of C11 experienced an upfield shift of 16ppm, which is consistent with the formation of the covalent aminal linkage at that position. Guanine N2 was identified as the site of anthramycin binding by default, as a large change in chemical shift would have been expected for GH8 if the aminal linkage had been formed at N7.

However, 2D NMR data gave a more detailed insight into this adduct (Krugh *et al.*, 1989). The duplex 5′-d(AT^{am}GCAT)·5′-d(ATGCAT) was used with one guanine sterically blocked, so that a single adduct was formed. Because anthramycin methyl ether undergoes rapid interconversion between 11(R) and 11(S) stereoisomers, it would be feasible that the exocyclic amino group during nucleophilic attack could attack either face, leading to two separate adducts. The fact that the side chain of anthramycin chain could extend in either direction along the duplex adds a further two elements to the equation. Interestingly, only one adduct out of the four variations is present. This is supported by strong NOE connectivities between H1a/H1b→A^{11}H2, indicating that the side chain is pointing in the 5′-direction of the modified strand. A weak NOE H11→H11a suggests that these neighbouring protons are trans to one another, while strong NOEs from H11→C^{4}H1′ and H11a→A^{11}H1′ demonstrate that H11 and H11a are each orientated towards a different strand. All these NOEs are consistent with the 11(S) configuration of anthramycin.

Pierce *et al.*, (1993) employed UvrABC nuclease to identify anthramycin and tomaymycin binding sites on modified DNA, which may have been overlooked by Hertzberg *et al.*, (1986) in their low sensitivity experiment. UvrABC nuclease digests in both directions along DNA, only stalling when it encounters a modified base. The 5′-AGA sequence was the preferred binding site of both ligands. However, the 5′- and 3′- flanking bases both influenced the affinities of these ligands for DNA. Anthramycin prefers 5′-A>G>T>C and 3′-A>G>T/C, whilst tomaymycin favours 5′-A>G/T>C and 3′-A/G>C/T. The preferred binding sites of anthramycin are therefore; 5′AGA, 5′AGG>5′GGA, 5′GGG, while tomaymycin favoured 5′AGA>5′GGC, 5′TGC, 5′AGC.

The origins of this observed specificity were explained by an X-ray crystal structure of the anthramycin–5′-d(CCAACGTTGG) adduct (Kopka *et al.*, 1994). The twist angle between purine-purine steps (A-A, A-G or G-G) was found to be smaller than average (24-37°), whilst in a G-A step it was slightly larger (31-42°). This twist enhances purine-purine ring stacking, which is advantageous energetically, even when their pyridine partners are destacked. Anthramycin and other PBDs recognize this twist in the base-pair sequence, which creates the perfect minor groove configuration for these ligands to fit into the 5′-TGG binding site. In the studied duplex, the twist angle at the T-G step is 51° but in the adduct this is lowered to 34°, demonstrating how anthramycin induces the fit of DNA around itself. The 5′-AT base pair creates an unobstructed, deep narrow minor groove in which the side chain can effortlessly lie.

Barkley *et al.*, (1986) looked at the tomaymycin-5′-d(ATGCAT)$_2$ adduct using various techniques. Evidence was presented for the existence of diastereomers *i.e.* 11S, 11aS and 11R, 11aS. This is in contrast to the findings of Graves *et al.*, (1985), where anthramycin adduct existed only as 11(S), 11a(S). To examine further whether tomaymycin would only bind as one diastereomer, a duplex with a single binding site 5′-d(CICGAATTCICG)$_2$ was used (Boyd *et al.*, 1990). Binding was found to take place at 5′-CGA, with C11 displaying 'S' stereochemistry and the side chain projecting in the 3′-direction. The production of >90% of a single tomaymycin species is therefore shown to be dependent on the flanking sequences. 5′-PyGPu led to one diastereomer whereas 5′-PyGPy resulted in two diastereomers, when equal proportions of 11(S)5′ existed. This is attributable to the lower affinity of tomaymycin for pyrimidines in the binding site (Pierce *et al.*, 1993). The most significant difference between these ligands, is that tomaymycin (unlike anthramycin) is unable to form stabilizing hydrogen-bonds (as demonstrated by NMR). These are likely to be an important factor in deciding the orientation of anthramycin orientation within the minor groove.

DSB-120 is a synthetic analogue of the naturally occurring PBDs (Fig. 15). The design exploits the knowledge gained by Boyd *et al.*, (1990), on the solely 11(S)3′ bis-tomaymycin adduct, where the side chain 'tails' effectively bind to DNA in a symmetric 'tail-to-tail' orientation. These PBD nuclei were attached together through a trimethylene bridge *via* O8 (Bose *et al.*, 1992). No phenolic hydroxy group is present at C-9 because anthramycin and sibiromycin (which both contain this functional group) are cardiotoxic, whilst neither tomaymycin or neothramycin are. Hurley and Thurston (1984) suggested that the cardiotoxicity maybe a result of oxidation or tautomerization of the C9OH to an orthoquinone imine. Each PBD nucleus is able alkylate the DNA from within the minor groove, the outcome of this reaction being an interstrand cross-link. Sequence-selectivity studies (Bose *et al.*, 1992), demonstrated that DSB-120 preferred a central AT base pair in the binding site between the GC bases, where the aminal linkages anchor the ligand in position. The DNA decamer 5′-

d($C^1I^2C^3G^4A^5T^6C^7I^8C^9G^{10}$) was chosen for an NMR study (Mountzouris *et al.*, 1994), because it contained a single target sequence in which DSB-120 could bind. A single interstrand cross-link was observed between G^4-N2 and G^{14}-N2. NOE data also suggested that a small perturbation in the helical structure existed at I^8, owing to its very weak internucleotide connectivities to C^7. This could occur if the bridge between the two alkylating moieties of DSB-120 was slightly too short, resulting in small positional rearrangements of the bases to accommodate the ligand more comfortably. The data also suggested that DSB-120 does not substantially sink itself into the minor groove. Secondary interactions were described by Jenkins *et al.*, (1994), which were in the form of two hydrogen-bonds originating from the aminal proton of DSB-120 to the adjacent 3′-A-N3 of the covalently modified G. In all, the ligand spanned 6bp pairs with the interstrand cross-link forming between four of these at 5′-GATC (Fig. 13).

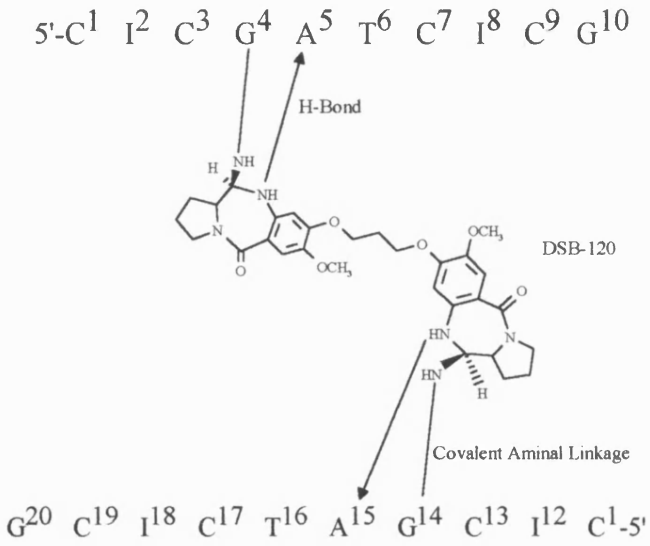


Figure 13 The interstrand cross-link formed by DSB-120.

Following the great improvement in cytotoxicity observed in the interstrand cross-linker DSB-120 compared to the monoalkylating parent compound, a C2/C2′*exo* methylene unsaturated analogue 'SJG-136' was synthesized (Gregson *et al.*, 1999), in an attempt to span a greater number of base-pairs and increase affinity. The unsaturated PBD structure was found to increase the melting temperature of calf thymus DNA by 33.6°C, whereas DSB-120 only produced a 15.1°C rise, illustrating the stabilizing effect of the C2/C2′-unsaturation. This ligand was found to be 9000x more potent than DSB-120 in human ovarian tumour cell lines.

The design and development of new cross-linkers could enhance sequence selectivity, adduct stability and binding affinity, whilst limiting unwanted side-effects seen in their monofunctional predecessors.

1.2.2 Cyclopropapyrroloindoles

(+)CC-1065 is an antitumour antibiotic which was first unearthed during routine screening of soil cultures by The Upjohn Company, Michigan (Hanka *et al.*, 1978). Produced by *Streptomyces zelenis*, (+)CC-1065 exhibits broad spectrum antimicrobial activity. However, it is its extreme potency as an antineoplastic agent against a wide variety of murine tumours which made it such a worthy pharmacological canditate for further research and development. This compound has also been isolated from *Streptomyces canulus*. Due to the therapeutic activity of (+)CC-1065 against experimental tumours in mice such as; P388 leukaemia, B16 melanoma and L1210 leukaemia (Martin *et al.*, 1978; Neil *et al.*, 1981), as well as promising results against human tumours *in vitro*, for example; breast, colorectal, ovarian and pancreatic carcinomas, neuroblastoma, adenocarcinoma of the lung and melanoma (Bhuyan *et al.*, 1981), (+)CC-1065 was accepted by the National Cancer Institute for clinical development as an antitumour agent. As an example of its potency, this drug produced 90% inhibition of L1210 cell growth *in vitro* while only at a concentration of 0·05ng ml^{-1}. Its nearest contender is Actinomycin D at 4·0ng ml^{-1}.

(+)CC-1065 is the parent compound in the cyclopropapyrroloindole family, so named after their common reactive moiety, which is responsible for alkylating DNA. The structure consists of three subunits; two identical pyrroloindole units (B and C) and the cyclopropapyrroloindole (subunit A) (Fig. 14).

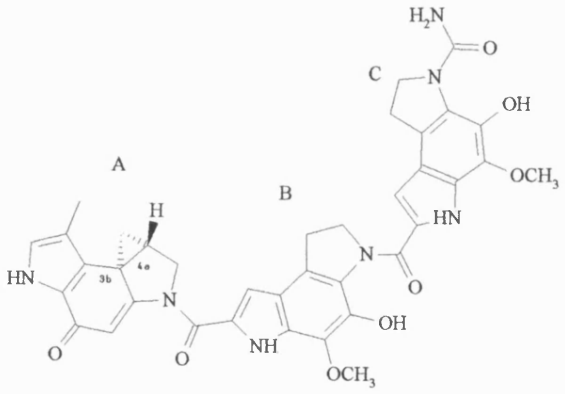


Figure 14 The molecular structure of (+)CC-1065 and its absolute stereochemistry at C3b and C4a.

In analogy with the DNA minor groove binders netropsin and distamycin, (+)CC-1065 has a right-handed twist along its axis relative to the out-of-plane bending of the amide linkers (Chidester *et al.*, 1981). The hydrophobic groups align perfectly along the concave edge of the molecule for complementary interactions with the minor groove, while the hydrophilic components protrude from the convex surface.

An abundance of research has taken place utilizing a vast range of techniques, to try to characterize the sequence-specificity and the mechanism of action of (+)CC-1065, in an effort to deduce why (+)CC-1065 exhibits such extreme biological potency. Early work looked at the effect of this ligand on L1210 leukaemia cells at the biochemical level (Li *et al.*, 1982). 50% inhibition of RNA synthesis required x10 the concentration of (+)CC-1065 to that needed to inhibit DNA synthesis. Protein synthesis was able to continue unaffected. Difference CD spectra showed (+)CC-1065 selectively bound to double stranded DNA. Lastly, only 1μg ml^{-1} of this ligand was required to inhibit the DNA polymerase activity by 70%, leading to the conclusion that (+)CC-1065 mediates its cytotoxic effects *via* DNA synthesis.

Further CD studies (Swenson *et al.*, 1982) established (+)CC-1065 had a preference for binding to p(dAT) rich oligomers, where p(dA)-p(dT) was favoured over p(dAT)-p(dAT). This interaction was very stable with no visible changes in the spectrum after one week's incubation, representing complex decomposition. Pre-incubation with netropsin severely inhibited (+)CC-1065 binding, whist (+)CC-1065 binding inhibited methylation and ethylation of G-N3, A-N3 and C-02. It was concluded that (+)CC-1065 covalently binds to DNA within the minor groove.

(+)CC-1065 is at least 100x more potent than adriamycin, as revealed by a human tumour cloning assay in which (+)CC-1065 caused ≥50% lethality at 0·1ng ml^{-1} within 1h against a variety of tumours (Bhuyan *et al.*, 1982). This great pharmacological profile was marred by a second observation that maximum inhibition of DNA synthesis did not occur until after 20 hours of exposure to (+)CC-1065. This was not due to cell death and hence loss of metabolic activity because RNA and protein synthesis continued. Instead, it suggested delayed expression of drug toxicity, correlating with the delayed death observed in (+)CC-1065-treated mice (Neil *et al.*, 1981). Non-tumour bearing mice and rabbits were given (+)CC-1065 by *iv.* and *ip.* administration (McGovren *at al.*, 1984), which produced delayed hepatotoxicity with some bone marrow depression at doses which were life-enhancing although not curative in tumour-bearing mice. Doses which did not result in delayed death were only borderline antineoplastic. Attempts to impede the hepatotoxicity by the administration of cytochrome P450 enzyme activators were to no avail, the toxicity being implemented by some other mechanism. These results precluded the development of (+)CC-1065 as a clinical antitumour agent by the National Cancer Institute but research has continued in the hope of producing an analogue with equal cytotoxic potency minus the hepatotoxicity.

To define precisely to which DNA base (+)CC-1065 binds, oligonucleotides containing [^{3}H]dA, [^{3}H]dG, [^{3}H]dC or [^{3}H]dT were treated with (+)CC-1065 to produce a drug-DNA adduct (Hurley *et al.*, 1984). This was heated, followed by butanol extraction of a

molecule containing the (+)CC-1065 chromophore, which proved to be covalently bound [^{3}H]dA-(+)CC-1065. NMR analysis established that the ^{13}C and ^{1}H NMR chemical shifts in this complex correlated well with those of 3-methyl adenine. The ^{13}C downfield shift of C4 was consistent with the cyclopropane ring opening and forming a covalent attachment to the N-3 of adenine. Further chemical shift changes of C-3b, C-7, C-8 and C-9 in the adduct were associated with the irreversible conversion of the cyclohexadienone to a phenol (Fig. 15). This base loss can be explained by the change in aromaticity of adenine, which effectively gains a positive charge.

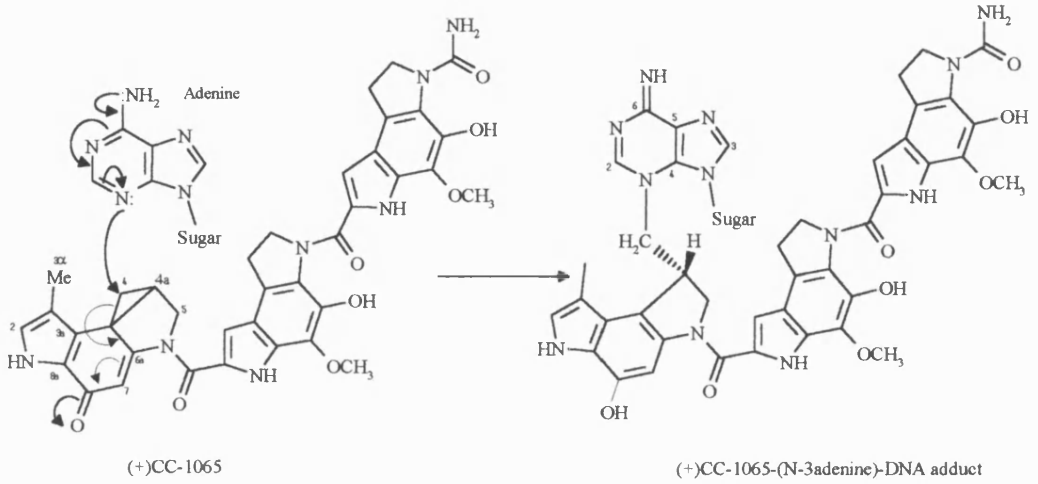


Figure 15 Mechanism of nucleophilic attack by the N3 of adenine upon the methylene (C4), electrophilic carbon of (+)CC-1065 cyclopropyl ring to form a (+)CC-1065-(N3-Adenine)DNA adduct

Hurley *et al.*, (1984) used (+)CC-1065 modified simian sarcoma virus (SVHO) restriction fragments, which were 5′-^{32}P end labelled, to probe the loss of this adduct from the DNA on heating. The (+)CC-1065-modified adenine is removed from the DNA strand *via* β-elimination, which involves cleavage of the glycosidic bond on heating, leaving this sugar without a base *i.e.* apurinic. Maxam-Gilbert sequencing, following piperidine cleavage of the DNA backbone, yielded two preferred binding sequences for (+)CC-1065; 5′-AAAAA* and 5′-PuNTTA*, where N is any base and * indicates the alkylated adenine. This sequence specificity is projected from the covalently modified adenine at the 3′-end, traversing five bases in the 5′-direction. Reynolds *et al.*, (1985) performed similar experiments using the SV40 early promoter element, which contained the AT rich TATA box. It was hoped (+)CC-1065 would bind preferentially here, explaining inhibition of DNA synthesis and thus the biological potency. In fact, the favoured binding site was 5′-AGTTA and not the TATA box. CPK models were also proposed for this adduct. Taking into account the polarity of drug binding, which indicates that the 'C' subunit projects in the 5′-direction from the covalently modified adenine, the DNA was used as a chiral reagent to detect the stereochemistry of C3b and C4a. (+)CC-1065 fits comfortably in the minor groove only when the stereochemistry at C4 is 'R'. The absolute

stereochemistry is therefore, C3b(R) C4a(S). Analysis of the stereo drawings also reveals a correlation between the subunits of (+)CC-1065 and the bases of the binding sequence. The cyclopropapyrroloindole moiety alkylates the 3′-A, the second benzodipyrrole subunit lies across a pair of highly conserved AA or TT bases (where 5′-TT is preferred over 5′-AA), and the third pyrroloindole 'C' group is adjacent to a less well conserved pair of bases 5′-PuN.

Additional confirmational data on the absolute stereochemistry at C3b and C4a in the (+)CC-1065–DNA adduct was acquired using the optical antipodes, U71-184 and U71-185, (Fig. 16) of a simple (+)CC-1065 analogue (Lee and Hurley, 1986). Only the biologically active U71-184 bound to DNA, with the absolute stereochemistry equivalent to the natural (+) enantiomer of (+)CC-1065. The cyclopropyl methylene carbon was then in close proximity to AN-3 to facilitate alkylation, instead of being directed away towards other base pairs.

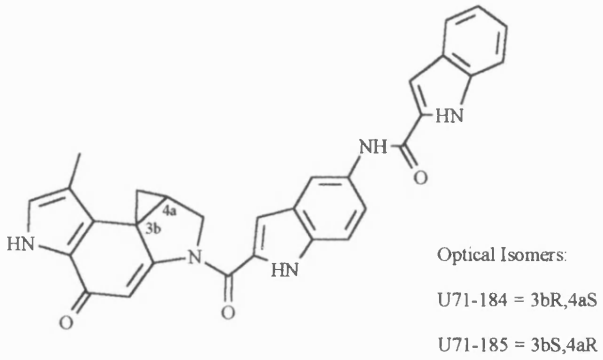


Figure 16 Molecular structure of the optical antipodes U71-184 and U71-185

Jacobsen *et al.*, (1986) examined detection and repair of this (+)CC-1065-DNA adduct within normal and Xeroderma pigmentosum (cells deficient in DNA lesion repair) human fibroblasts. It was found that this adduct was the source of a defect between the excision and ligation processes. Recognition, possibly due to the positive charge on adenine, and excision of covalent (+)CC-1065-modified DNA, are the first steps in DNA repair. Strand breaks are induced by the excision enzyme and the adduct is removed. The presence of a strand break activates the chromatin-associated enzyme, p(ADP-ribose) polymerase, which utilizes NAD as its substrate for this repair. However, when the incidence of strand breaks is high, the rate of poly(ADP-ribose) synthesis exceeds that of NAD synthesis, leading to a depletion in the cellular NAD pool. This in turn leads to incomplete repair by the enzyme, and the inability of the cell to maintain normal energy metabolism resulting in apoptotic cell death.

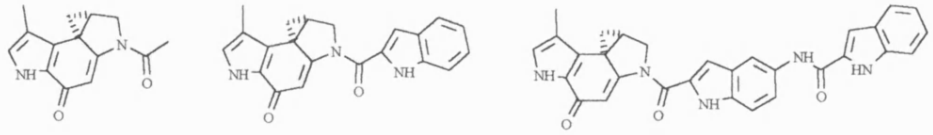
Hurley and Needham-VanDevanter (1986) analysed the binding of (+)CC-1065 to the minor groove. The covalent event is only initiated after noncovalent binding interactions have optimized sequence specificity and positioning of the ligand within the minor groove,

postulating that adduct formation is biphasic. These noncovalent interactions include stabilizing electrostatic forces, hydrophobic interactions between the concave surface of (+)CC-1065 and particularly adenine H-2 protons, which strengthen the binding affinity, plus many van der Waals contacts between (+)CC-1065 and the floor of the narrow minor groove. Van der Waals forces drive the steric manoeuvrings to maximize the affinity for the ligand. As previously discussed with other minor groove binders, hydrogen-bonding does not have a role in sequence specificity. Its premier function is stabilization of the drug-DNA complex.

DNase I footprinting was utilized by Hurley *et al.*, (1987) to examine what effect (+)CC-1065 alkylation of DNA had on local helical structure. Mild inhibition of restriction enzyme cleavage projects as far as eleven base-pairs to the 5′-side of the covalently modified adenine. More pronounced enzyme inhibition occurs on the non-alkylated strand (-), over a 12 base-pair region, which is particularly notable opposite the 'A' subunit and for at least 5 bases to the 5′-side along this (-) strand. This reflects the (+)CC-1065 induced helical conformation changes specifically adjacent to the 'B' and 'C' subunits, and to the 5′-side of this on the opposite modified (+) strand. (+)CC-1065 produces an asymmetric effect on DNA conformation.

Warpehoski *et al.*, (1988) synthesized a series of (+)CC-1065 analogues to explore structure-activity relationships. Some important stuctural attributes became apparent during this investigation. Firstly, an acyl group substituted on the CPI pyrrolidine N enhanced electrophilic reactivity of the CPI unit and consequently acid-catalyzed nucleophic attack, which was thought to be important for alkylation. Secondly, DNA binding increased when the amide-linked indole chain of the central and right hand segments increased in length. This resulted in intensified hydrophobic interactions which promoted van der Waals forces and, hence, drug-DNA adduct formation. Thirdly, the ring size of the 'B' subunit can limit the conformations of the ligand within the minor groove. A five-membered ring permits binding to the helix. O-catechol substituents are required for the tight complexation of (+)CC-1065 within the minor groove (Warpehoski and Bradford, 1988). The ethylene bridge moieties of the B and C subunits, are thought to be responsible for the unusual delayed hepatotoxicity exhibited by (+)CC-1065 (McGovren *et al.*, 1984; Warpehoski and Bradford, 1988).

Continuing their review (Hurley and Needham-VanDevanter, 1986) of non-covalent binding interactions, Hurley *et al.*, (1988a) prepared four analogues of (+)CC-1065 (Fig.17). Their non-covalent versus covalent interactions were evaluated relative to sequence-specificity and biological potency, in comparison with (+)CC-1065.



(+) A (+) AB (+) ABC

Figure 17 Structures of the (+)CC-1065 synthetic analogues.

Their findings stated that biological potency was linked to alkylation of DNA. This was established using an analogue identical to (+)ABC minus the cyclopropyl ring, which resulted in no binding to DNA. Unlike distamycin and netropsin, the non-covalent interactions involved in binding were not strong enough to form a drug-DNA complex detectable by footprinting. The B and C subunits were found to increase the rate of alkylation at specific adenines. However, (+)A, (+)AB and (+)ABC demonstrated equivalent sequence-specificity in adduct formation, implying that it was the DNA sequence which was responsible for the reactivity of particular adenines towards the CPI subunit of (+)CC-1065. The B and C subunits were thought to take part in stabilizing non-covalent bonding, as well as fine-tuning sequence selectivity prior to alkylation. It is possible that the 5′-flanking sequence shows some conformational flexibility, which allows the target adenine in a particular locale to approach the electrophilic centre and form a covalent bond *via* nucleophilic attack. This results in a sequence-dependent, Lewis acid-catalyzed DNA alkylation, requiring autocatalytic activation of the alkylation step. This is achieved through carbonyl complexation or protonation *via* a strategically located phosphate in the DNA backbone, 2bp removed from the alkylation site in the 5′-direction. Hurley *et al.*, (1990) suggested a kinetic binding scheme to support this theory:

$$\text{drug} + \text{DNA} \overset{K_b}{\rightleftharpoons} (\text{drug-DNA})_{\text{noncov}} \overset{K_r}{\rightarrow} (\text{drug-DNA})_{\text{cov}}$$

A high binding affinity (K_b) indicates high sequence-specificity of the drug or that the activation energy for reaction is very favourable for the sequence. If K_r was not sequence-dependent, covalent reaction would take place at sites where drug dissociation was the slowest, or where K_b was highest if K_r was slow. (+)CC-1065 has a very high K_b, indicating a high degree of noncovalent binding, but adduct formation is constrained by the sequence dependent K_r.

Lin *et al.*, (1991) analysed (+)CC-1065-5′-d(CGCGGAGTTA*GG) by NMR using ^{17}O-labelled water and phosphate. This led to the incorporation of two water molecules into the alkylation reaction pathway. One bridges the 8-phenolic proton of the (+)CC-1065 CPI subunit to the anionic oxygen of the phosphate, two bases to the 3′-side of T constituting the covalently modified base pair. A second ordered water molecule was hydrogen-bonded to the alkylated adenine. Due to prolonged dwelling times of these water molecules, it was proposed that they were involved in the mechanism of covalent bond formation, their role possibly being acid catalytic activation of the CPI in its reaction with N3 of adenine.

In sharp contrast to Upjohn and Hurley's suggestion that (+)CC-1065 alkylation of DNA was a sequence-dependent event, in which the non-covalent binding abilities of the drug bore minimal importance to its sequence-selectivity, Boger and co-workers have presented insuperable evidence from a variety of experiments that this is not the case. Many (+)CC-1065 analogues were synthesized, one of which substituted a benzindole group for the pyrroloindole in the CPI head unit, to form (+)CBI-CDPI$_2$ (Fig. 18) (Boger and Ishizaki, 1990). This compound exhibited greater cytotoxic potency than either (+)CC-1065 or (+)CPI-CDPI$_2$ (Fig. 18), proving that biological activity is not dependent on electrophilic reactivity. Examination of the rates of solvolysis demonstrated the relationship between electrophile reactivity and rate of DNA alkylation was in fact inverse and not a direct one. A further set of (+)CC-1065 analogues, CPI-CDPI$_1$, CPI-CDPI$_2$, CPI-CDPI$_3$ and (+)N-BOC-CPI (Fig. 18) in both their (+) and (-) enantiomeric forms were used in a structure-cytotoxicity study (Boger *et al.*, 1990) to elucidate the role of the central and right-hand subunit of (+)CC-1065. (+)CPI-CDPI$_2$ and (+)CC-1065 displayed an indistinguishable alkylation profile. This demonstrated that removal of hydroxy and methoxy substituents had only a small impact on the binding affinity and no impact on the AT-rich binding selectivity. (+)CPI-CDPI$_1$ and (+)CPI-CDPI$_3$ were marginally weaker alkylators than was (+)CC-1065, whilst (+)N-BOC-CPI required a concentration x10^7 greater to produce alkylation comparable to (+)CC-1065. The unnatural (-) enantiomers required all three subunits for effective alkylation, which, at low concentration, was at different sites to those occupied by (+)CC-1065. This evidence clearly supports a binding model where subunits B and C play a large functional role in sequence-selectivity, through non-covalent binding preferentially in the AT-rich minor groove.

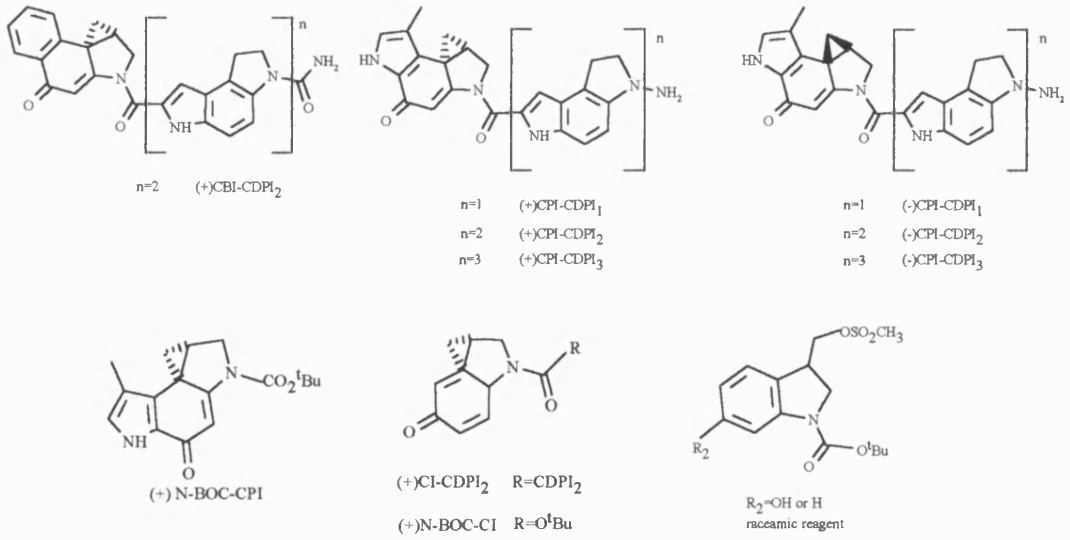


Figure 18 Structures of the various (+)CC-1065 analogues Boger *et al.*, have used in their studies.

(±)N-BOC-CI (Fig. 18) is an authentic version of the CPI alkylation subunit which binds to DNA indiscriminately at any adenine (Boger *et al.*, 1991a). Addition of a $CDPI_2$ subunit yields $CI\text{-}CDPI_2$, which has the ability to alkylate specific adenines as a consequence of selective noncovalent interactions (Boger *et al.*, 1991b). Removal of the cyclopropane ring results in a racemic reagent (Fig. 18) which can still alkylate adenine. This suggests the cyclopropyl ring may not be a fundamental requirement for covalent binding although if R_2 = OH, the cyclopropane could still form as an intermediate. The solvolytic reactivity versus cytotoxic potency was compared for the compounds (+)$CPI\text{-}CDPI_2$, (+)$CI\text{-}CDPI_2$ and (+)$CBI\text{-}CDPI_2$. Using their intensity of alkylation as a measure, this experiment showed that an inverse relationship existed between electrophilic reactivity and cytotoxic potency: (+)$CBI\text{-}CDPI_2$ > (+)$CPI\text{-}CDPI_2$ > (+)$CI\text{-}CDPI_2$. (+)$CBI\text{-}CDPI_2$, the most stable and least electrophilicly reactive agent produced the most DNA alkylation, where the sequence preference followed the order: 5′-d(AAA)>d(TTA)>d(TAA)≥d(ATA)-3′. The fourth base in the 5′-direction was also A or T.

Despite the large modification of the alkylating subunit which the substitution of a benzene for a pyrrole ring incurs, it is proven to be well-tolerated with a potency and stability x4 to that of corresponding CPI analogues (Boger *et al.*, 1995). Boger and Mesini (1995) noted that nucleophilic addition was dictated by stereoelectronic preferences. A-N3 alkylation took place at the least substituted cyclopropane carbon. Boger and Johnson (1995) presented a last piece of evidence to prove beyond doubt that non-covalent binding of the B and C subunits was responsible for the sequence-specificity observed on (+)CC-1065 adduct formation. A comparison was made of the observed binding directions of some natural and unnatural enantiomers. Analogues were synthesized of the duocarmycins (DSA), which contain the same pharmacophore as (+)CC-1065. The ester of duocarmycin allows substitution of non-covalently binding subunits on either side of the alkylating moiety (Fig. 19). Like (+)CC-1065, (+)$DSA\text{-}CDPI_2$ alkylated the most 3′ adenine of the sequence 5′AAAAA*. (-)$DSA\text{-}CDPI_2$ alkylated the same position as (-)CC-1065 in the sequence 5′AA*AAA (where * denotes covalent modification).

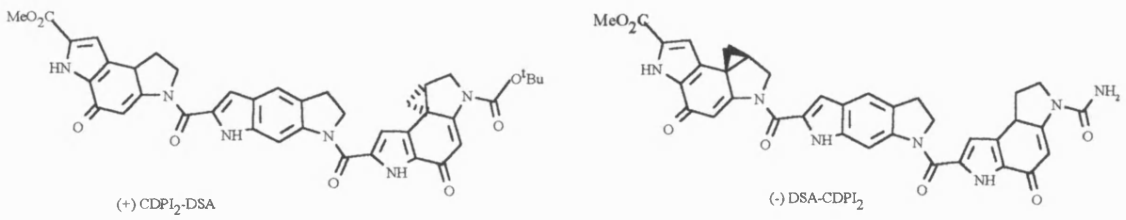
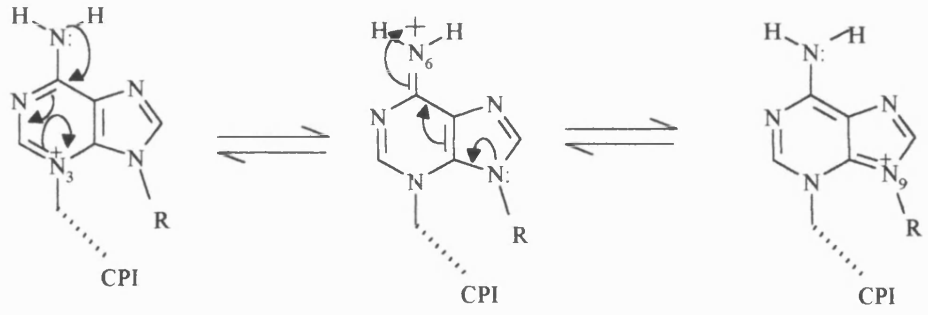


Figure 19 The molecular structures of (+)$CDPI_2$-DSA and (-)$DSA\text{-}CDPI_2$.

Due to the reversed nature of (+)$CDPI_2$-DSA, although binding takes place within the same adenine sequence, alkylation now occurs at the same site as (-)CC-1065 because it now extends along the groove in the opposite direction. The B & C subunits of (+)$CDPI_2$-DSA extend along the groove in the opposite direction to (+)CC-1065. The reversal of enantiomeric alkylation selectivity resulting from the reversal in ligand orientation confirms that the same underlying recognition features, *i.e.* non-covalent interactions, are in action for both natural(+) and unnatural(-) enantiomers.

Lin and Hurley (1990) used NMR as a tool to elucidate the tautomeric form of the covalently modified adenine. A large downfield shift of a resonance in the ^{15}N NMR spectrum of the adduct was attributed to the formation of a positive charge and partial double bond character between N6 and C6 in the alkylated adenine. To ascertain whether the positive charge was situated at N3, N6 or N9, quantum mechanical calculations were performed on 3, 9-dimethyladenine. This concluded the charge was delocalized over the entire adenine molecule (Fig. 20). Due to N6 being predominantly doubly protonated as the 6-amino form, the bond distance between C6 and N6 is computed to have shortened.



Major tautomeric form

Figure 20 Tautomers of adenine in its amino form, resulting from covalent (+)CC-1065-DNA adduct formation.

Powers and Gorenstein (1990) studied the (+)CPI-$CDPI_2$-5′-$d(C^1G^2C^3T^4T^5A^6A^7G^8C^9G^{10})_2$ adduct by 2D NMR. Only one major adduct resulted from the CPI-$CDPI_2$ (analogous structure to (+)CC-1065, minus -OCH_3 and -OH on the B and C subunits) reaction with DNA. A^{17} was the alkylation site with the C subunit being orientated towards the 5′-terminus of the modified strand. NOESY-generated distance restraints were used to produce a refined molecular model. CPI-$CDPI_2$ binding within the minor groove was found to distort the β-helical structure of DNA. A deviation of 60° from the helix axis at A^{17} produced a bent DNA structure with a wider minor groove.

Scahill *et al.*, (1990) studied the (+)CC-1065-5′-d($C^1G^2A^3T^4T^5A^6G^7C^8$) adduct by NMR. The COSY spectrum displays large downfield shifts of the cyclopropyl protons into the methylene region, signalling that the ligand is covalently bound. Intermolecular NOEs with minor goove H2 and H1′ protons verified that (+)CC-1065 was alkylating A^6. ^{13}C NMR data also revealed that C4a had a chemical shift of δ 40·1ppm. This is indicative of a methine group, resulting from cyclopropane ring opening, which is adjacent to two methylene C (C4 and C5) and an aromatic ring. This concluded that the nucleophilic addition progressed as in Figure 21a. If the mechanism of reaction had been as in Figure 21b, then the neighbouring N-3 would have sent the methine resonance to a lower field (>50ppm).

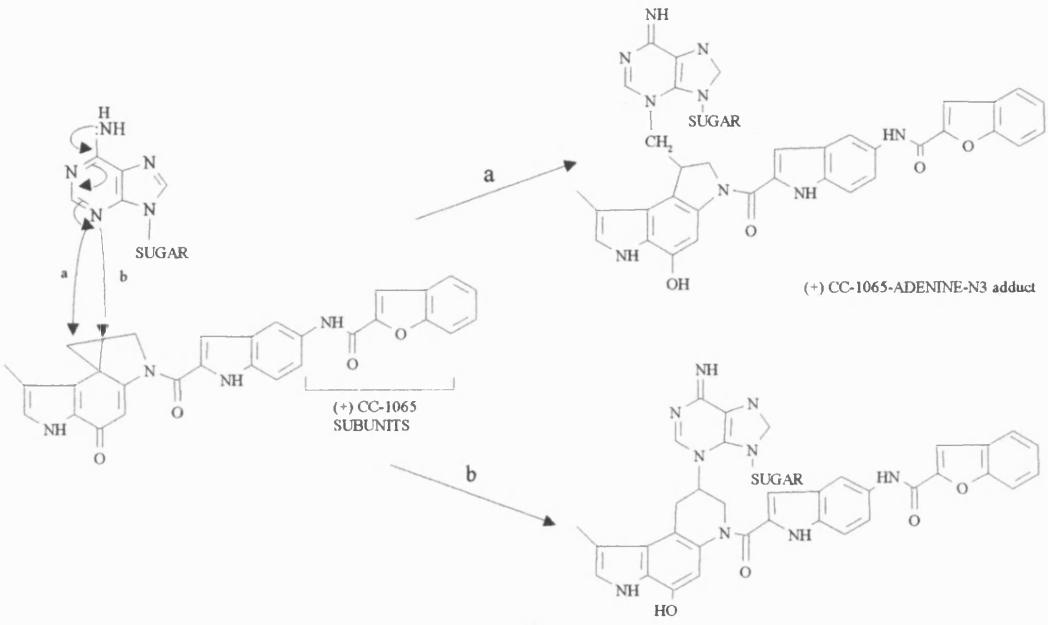


Figure 21 Diagram showing the two possible reaction pathways which could result from adenine-N3 nucleophilic attack on (+)CC-1065 (Scahill *et al.*, 1990).

(+)CC-1065 is responsible for the blockage of a number of enzymes involved in DNA synthesis. The progression of Klenow fragment *E.coli* DNA polymerase and T4 polymerase along DNA is obstructed on encountering the (+)CC-1065-DNA adduct (Sun and Hurley, 1992a). This blockage may result in the arrest of DNA synthesis, although some evidence exists that the presence of (+)CC-1065 may lead to the misincorporation of dG and dA bases, indicating that this ligand has mutagenic properties. (+)CC-1065 also inhibits T4 ligase and helicase II, which are responsible for unwinding DNA prior to DNA synthesis and repair. It achieves this by stiffening and bending the DNA by approximately 17° in the direction of the minor groove, resulting in a winding effect within the adduct site. The reduction of a helix turn by one base pair changes the topology of the binding site. Winding effects maybe related to prevention of adduct excision and the expression of delayed lethality.

M^{c}Hugh *et al.*, (1994) compared the reaction of (+)CC-1065 with purified and cellular DNA. The structures of the resulting adducts were essentially the same, except that higher concentrations of drug were required for cellular DNA to produce the same effect as with naked DNA. Using spectroscopic techniques to measure the rate of covalent additon of (+)CC-1065 to DNA, Warpehoski and Harper (1995) established that the speed of alkylation was equivalent to that catalysed by an enzyme. Gunz and Naegeli (1996) developed a translocation assay to analyse the contribution of non-covalent interactions to binding site recognition in DNA. Addition of (+)CC-1065 to DNA at saturation concentration produced a complex in which only 30% of the bound (+)CC-1065 was attached covalently. However, the non-covalently bound (+)CC-1065 was resistant to removal by dialysis and precipitation techniques. On incubation with DNA fragments containing a preferred binding site, these noncovalently bound (+)CC-1065 molecules dissociated and translocated to the alkylation site. This molecular translocation or 'hopping' mechanism along DNA in search of a specific binding sequence, illustrates the concept of 'biphasic binding'.

Adozelesin (U73-975) is a second-generation synthetic analogue of (+)CC-1065. Like the parent compound, adozelesin consists of three subunits; a cyclopropapyrroloindole (A), an indole (B) and a benzofuran (C), all linked together *via* amide bridges (Fig. 22). Closely resembling the structure of (+)CC-1065, adozelesin can mimic the pitch of β-helical DNA helix minor groove perfectly, forming a covalent DNA adduct *via* N3 of adenine. Tayeb *et al.*, (1996) established, using CD spectroscopy, that adozelesin exhibits a left-handed conformation at low and room temperatures but, at high temperature or on reaction with DNA, it reverts to a right handed conformation.

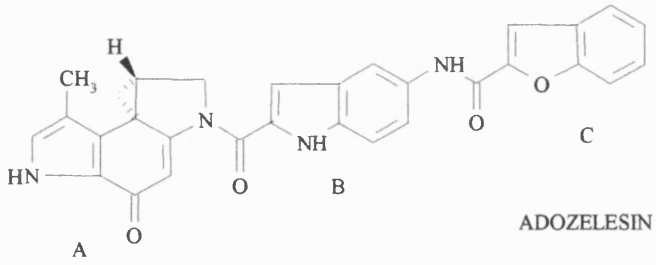


Figure 22 Structure of adozelesin

Lee and Gibson (1991) noted that adozelesin at concentrations between 20-60pM, could produce a 1 log cell kill in 6 human tumour cell lines. This makes it more potent than (+)CC-1065, with very broad spectrum activity. Adozelesin has displayed a greater potency than cisplatin, 5-fluorouracil and doxorubicin in human gynaecological cell lines (Ngugen *et al.*, 1992). Sequencing techniques were used to probe the consensus sequence of adozelesin (Weiland and Dooley, 1991). Blockage of the progression of bacteriophage DNA polymerases

along the template strand occurred on encountering the adozelesin adduct, specifically at the nucleotide alongside the covalently modified base. Two preferred binding sequences were identified, p(dA) and 5′-(T/A)(T/A)T-A*(C/G)G. Yoon and Lee (1998) used thermally induced strand cleavage assays to elucidate the consensus sequences for adozelesin binding. Their findings largely agreed with these of Weiland and Dooley (1991). The preferred binding sequences were 5′-(A/T)(A/T)A*, 5′-(A/T)(G/C)(A/T)A* and another more rare sequence 5′-(A/T)(A/T)CA*, which has a cytosine adjacent to the modified adenine in the 5′-direction. Tritium-labelled adozelesin was detected at high concentration on cellular nuclear DNA at 5′-TTAGGG.

Bhuyan *et al.*, (1992) compared the cytotoxicity of adozelesin in human and rodent cells. The ligand was shown to be ×10 more cytotoxic against the human ovarian carcinoma than against hamster ovarian carcinoma and lung or mice melanomas, demonstrating a lack of correlation between cytotoxicity and the extent of DNA alkylation. This could be interpreted as either the sites of alkylation varying in proportion between species influencing the degree of cytotoxicity or a difference in the capacity of the cell type to repair covalently modified DNA. Inhibition of cell growth occurrred at doses which only marginally inhibited DNA synthesis. Bhuyan *et al.*, (1992) examined the effect of adozelesin on the mitotic cell cycle. Initially, progress through S phase (DNA synthesis and replication) was slowed, followed by block of G2 phase (cell metabolism and where performs functions as an organism prior to cell division, *i.e.* mitosis). This was superceded by lysis and cell death some hours later. The precise mechanism by which adozelesin inhibits DNA replication was investigated using SV40-infected African green monkey kidney cells (Cobuzzi *et al.*, 1996). Low adozelesin concentrations were found to inhibit the initiation of SV40 replication, while at higher doses elongation of nascent DNA chains was also inhibited.

Baraldi *et al.*, (1997) synthesized analogues of adozelesin and U71-184, which incorporated a pyrazole into the CPI headunit (Fig. 22). These compounds demonstrated AT selectivity on binding upstream of the human oestrogen receptor, rather than the Ha-ras oncogene which is GC rich. This analogue (Fig. 23), was found to be considerably less potent *in vitro* than *in vivo*, where it exceded the potency of adozelesin. This data agrees with Boger's statement that modification of the alkylating motif was well tolerated, even increasing cytotoxic potency.

On the grounds of the greater cytotoxic potency, solubility and stability in aqueous formulations of adozelesin in comparison to (+)CC-1065, plus its freedom in animal models not only from hepatotoxicity but also to prolong life and inhibit tumour growth, adozelesin was

chosen as a candidate for clinical trial. A phase I clinical trial was conducted by Shamadas *et al.*, (1994) to assess the effect of this drug on patients with drug-refractory or solid tumours.

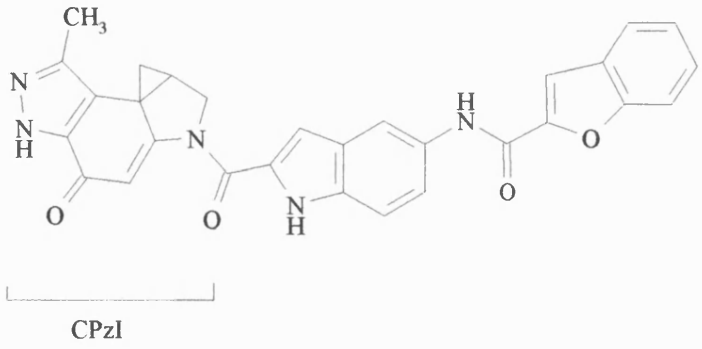
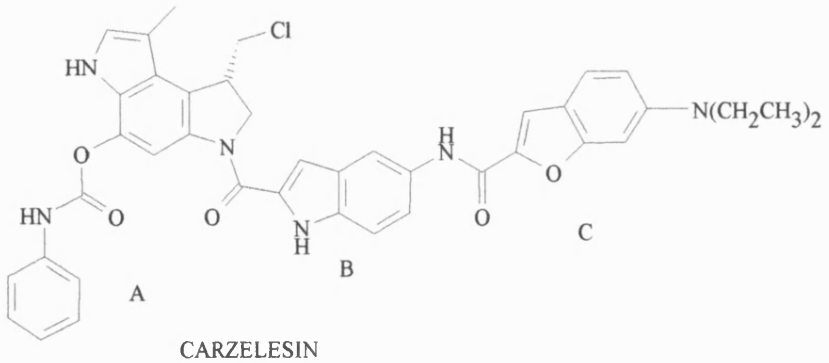


Figure 23 Structure of the adozelesin analogue which incorporates the CPzI subunit (Baraldi *et al.*, 1997)

Adozelesin proved to be generally well tolerated over a range of doses. Drug limiting toxicities were quite serious, however, myelosuppression being the main one with thrombocytopenia (very low platelet count) being more pronounced than leukopenia. A similar trial by Burris *et al.*, (1997) on patients with solid tumours reached the same conclusion that, although no objective antitumour activity was apparent, the toxicity and *in vitro* activity profile justified its progression to Phase II clinical trial. Cristofanilli *et al.*, (1998) clinically evaluated the therapeutic efficacy of adozelesin against incurable metastic breast cancer. With the dosage and schedule described, adozelesin showed only marginal efficacy with some patients actually showing progression in their disease. This trial was aborted in favour of more effective drugs. It is not known why adozelesin, which is extremely potent *in vitro* appears not to be *in vivo*; possibly its biological half-life is too short. However, this drug seems to exhibit the opposite toxicity profile *in vitro* and *in vivo* to the CPzI analogue described by Baraldi *et al.*, (1997).

Carzelesin (U-80244) is also a second generation analogue of (+)CC-1065. It exists as an inactive prodrug, requiring chemical or enzymatic activation of the relatively unreactive chloromethyl precursor to form the cyclopropyl moiety. Activation of carzelesin proceeds by hydrolysis of the phenylurethane substituent to a chloromethyl phenol, followed by a ring-closing step to produce a cyclopropal keto compound ready to alkylate DNA. The synthesis of a CPI prodrug was part of an investigation to ascertain whether the rate of alkylation could be regulated and, hence, to see if this in turn increased antitumour efficacy. Examination of the structure of carzelesin shows the usual CPI keto moiety now exists as a phenolic one, which in the prodrug form is derivatized with a carbamate. The resulting compound has marginally lower potency but equivalent efficacy *in vivo* to the ordinary cyclopropyl derivatives. The benzofuran 'C' ring was also modified with a diethylamino substituent to improve drug efficacy (Fig 24).



CARZELESIN

Figure 24 Chemical structure of carzelesin.

Li *et al.*, (1992) assessed the performance of carzelesin as a prodrug. The data from cell culture experiments suggested that carzelesin had enough time to circulate throughout the medium before being transformed to the active compound U-76074 containing the cyclopropane. Incubation of carzelesin with L1210 leukaemia cells showed that this compound was approximately x1500 less cytotoxic than U-76074 after 1h. After 24h, this difference had shrunk to x2-10, relative to the conversion of carzelesin into its biologically active form *i.e.* U-76074. Carzelesin induces only a small change in the CD difference spectrum when in the presence of DNA, unlike U-76074. This suggests carzelesin has not formed a stable covalent DNA adduct but the ligand containing the active cyclopropyl group has, in an analogous manner to (+)CC-1065. The prodrug carzelesin demonstrated a broad spectrum of antitumor activity, with more survivors which had longer or complete remissions than adozelesin. This is probably linked to the improved pharmacokinetic and distribution properties available with the prodrug. The successful pharmacological profle of carzelesin secured its position in clinical trials. Phase I trials concluded that myelosuppression, *i.e.* neutropenia and thrombocytopenia, were the dose limiting toxicities of carzelesin, occurring after only one exposure (Wolft *et al.*, 1996). The origin of this toxicity has been investigated in mice (Filippini *et al.*, 1997). After an initial drop in mature bone marrow cells post carzelesin administration, cell proliferation began again within 24h and, within 14 days, the bone marrow was repopulated with all cell lineages. This suggested that, with an improved dosing regimen which allowed replenishment of bone marrow cell stocks after drug administration, more cycles of drug therapy would be possible without cumulative myelotoxicity.

Phase II clinical studies (Sorio *et al.*, 1998) on patients with non-Hodgkin's lymphoma demonstrated the same significant haematological toxicity profile as seen with Phase I trials. At these doses of carzelesin prescribed, only one partial response in seventeen patients was noted. This seems to mimic the problem of lack of clinical effectiveness, as already seen with adozelesin in humans. Indeed, in a comparative study of carzelesin administration to rats, mice and humans, van Tellingen *et al.*, (1998) showed that carzelesin was tolerated by approximately

5 and 66x more in rats and mice respectively than humans, possibly compromising its success as an antineoplastic agent.

Bizelesin (U77-779) is another second generation (+)CC-1065 analogue. It is unique in having two CPI headunits, one at each end of the molecule, which confer bifunctional alkylation activity, resulting in the production of DNA interstrand cross-links. These CPI subunits are each attached to an indole, which are subsequently joined tail to tail *via* a ureylene linkage (Mitchell *et al.*, 1991) (Fig 25). Intramolecular rearrangement within bizelesin results in conversion of the chloromethyl moieties, into the cyclopropane reactive species (U77-809) which alkylate DNA in a mode analogous to (+)CC-1065.

Bizelesin

Figure 25 Structure of bizelesin.

The length of the bizelesin molecule means that the interstrand cross-link formed within the minor groove spans 6 or 7 nucleotides, depending on the composition of the intervening base sequence (see later explanation). 5′-•TAATTA* (where * denotes covalent modification of the adenine on this strand and • denotes covalent modification of the adenine to which this thymine is base-paired to on the complementary strand) is the bizelesin cross-link consensus sequence (Mitchell *et al.*, 1991). Bizelesin has demonstrated good antitumour efficacy both *in vitro* and *in vivo*, even against drug-resistant murine tumours such as pancreatic O2 carcinoma and colon 38 carcinoma. At pM concentrations it is 2-30x more potent than adozelesin (Lee and Gibson, 1991), without the hepatotoxicity of (+)CC-1065.

Sequence-specificity of bizelesin was found to be susceptible to the pH conditions of the experiment (Lee and Gibson, 1993). At the highly alkaline concentrations used to extract and examine damaged DNA, the bizelesin interstrand cross-link was unstable with adduct reversal taking place. This effect was not noted when bizelesin monoalkylated DNA. However, at low pH and high temperature, depurination (to which the interstrand cross-link was resistant) took place probably after the DNA melted to single stranded. Interstrand cross-links occur at 5′-•TAATTA*, which contains the classic (+)CC-1065 consensus sequence, but they also arise less frequently at 5′-•TCCGCG* and 5′-TTCCGC (Sun and Hurley, 1993). This alkylation of GN3 most probably occurs after the covalent binding of a bizelesin (+)CPI headunit to an adenine, which leads to the remaining (+)CPI subunit being placed in the close vicinity of the

GN3 6 base-pairs away where the second alkylation takes place. The ultimate formation of the interstrand cross-linked product results primarily from the trapping out of a thermodynamically more stable product, which maybe favoured by minimal distortion of the helix resulting in a low binding energy penalty or from a kinetically favourable reaction at the first site. Cross-linking involving a guanine was also demonstrated (Lee *et al.*, 1996) in an experiment using the consensus sequence 5′-TAATTN, in which N was G, C or T. Bizelesin did not form interstrand cross-links when N was a pyrimidine base. Lee *et al.*, (1994) compared bizelesin binding to genomic and naked DNA. In both cases, binding was sequence-specific but there was a difference in the consensus sequences, probably due to the presence of histones and other nuclear proteins in genomic DNA. Monoalkylation occurred at the (+)CC-1065 consensus sequence 5′-(A/T)(A/T)A*, whilst the preferred cross-link sites were 5′-•T•TTTTTA* (which is intrinsically bent), 5′-•T•TTATCA* and 5′-G•TACTAA*, where one GC base-pair was tolerated within the binding site.

NMR studies of bizelesin binding to 5′-•TAATTA* (Seaman and Hurley, 1993), noted that the central base-pairs within the cross-link were trapped in two different conformations. In 40% of the species the central base-pairs were open, whilst in 60% of species they were Hoogstein base paired. They also observed a loss of 'A' tract character such as propellor twisting on bizelesin binding to 5′-•TTTTTA*.

Thompson *et al.*, (1995) examined the role of the central GC base-pair within the 7 base-pair cross-link of 5′-CG•TTAGTTA*CG, *via* 2D NMR and molecular modelling. Bizelesin is too short to form a cross-link spanning 7 base-pairs without distortion of the duplex. NOE connectivities revealed that this adduct existed as a single species, retaining Watson-Crick base-pairing, and was orientated edge on in the minor groove, alkylating A^9 and A^{20}. This particular interstrand cross-link occurs even though there is an adenine available for binding at the sixth base in the sequence, suggesting there is a structural feature responsible for the 7 base-pair preference. Distortion of this sequence mainly takes place around each alkylated adenine, which are pulled down towards each other by bizelesin to decrease the distance it has to transverse by half a base-pair, creating an adduct with a high degree of propellor twisting throughout the binding site. The role of the G was unveiled by contrasting this adduct with the one formed in the presence of an inosine, in which bizelesin only traversed 6 base-pairs. It was concluded from these data that G-NH_2 hydrogen-bonded with the C=O of the ureylene linker, facilitating distortion of the helix so bizelesin could form a stable interstrand cross-link.

Thompson and Hurley (1995) have also investigated the nature of the bizelesin interstrand cross-link within the 'A' tract sequence 5′-CG•TTTTTA*CG, using the same techniques. A single adduct was formed which crossed 6 base-pairs. The previously described

base-pair open conformation within the centre of the duplex, was not observed due to the long half-lives of AT base-pairs. The ureylene linkage of bizelesin oscillated freely between two conformations, with no hydrogen-bonding taking place in either. NOE data demonstrated that the helix had lost its 'A' tract character *i.e.* propellor twisting and junction sites at either end of the 'A' tract and, instead, the adduct features were representative of straight β-form DNA. On formation of the interstrand cross-link, bizelesin trapped each covalently modified adenine in its β-form. This inhibited the propellor twist of the flanking bases and, reduced the length of the 'A' tract from five bases to two, which is not enough to induce propellor twisting, so bizelesin binding produced straight DNA.

A quantitative evaluation of bizelesin adducts formed in genomic and SV40 DNA, showed that in contrast to cisplatin which produced $1{\cdot}2 \times 10^6$ lesions per cell, and (+)CC-1065 which yielded $\sim 10^4$ lesions per cell, bizelesin required only 10^2 lesions per cell to reduce cell growth by 90%, demonstrating its extreme potency (Woynarowski *et al.*, 1995). Bizelesin interstrand cross-links, however, do take three times longer to form than ones resulting from monofunctional alkylation. Even though bizelesin produces fewer lesions, they appear to be more specific, as one would expect for an agent requiring a longer consensus sequence for bifunctional alkylation. Drug-induced damage occurred frequently at the origin of replication and topoisomerase II cleavage sites, hinting at the mode of antineoplastic activity of bizelesin.

Seaman *et al.*, (1996) looked at the binding of bizelesin to 5′-TAATTA, when the ligand was separated into two components by removal of the ureylene linker. It was expected that they would bind to the consensus sequence in the same manner as bizelesin, rendering a symmetrical adduct. In fact, the disconnected CPI-I subunits alkylated 5′-TAA^{2}TTA^1 (numbers ^{1,2} denote order of alkylation) on the same strand, with retention of β-helical structure. Gel electrophoresis and molecular dynamics studies of the CPI-I 5′-TAA^{2}TTA^1 and 5′-TAA^{2}AAA^1 bisadducts suggested that the first alkylation produced a distortion in the binding sequence at the 5′-TA junction site, making it unfavourable for symmetrical bisalkylation. Helical deformation was retained on the second CPI-I binding. To test whether the 'head to tail' binding mode was a result of 5′-TAA^{2}TTA displaying increased reactivity by co-operative bending induced by the first alkylation, or whether the 5′-TAATTA^1 was made less reactive by distortion of the minor groove on monoalkylation, a 5′-TAAAAA consensus sequence underwent bisalkylation. The resulting bisadduct 5′-TAA^{2}AAA^1 and 5′-TAAAA^{1}A monoadduct, demonstrated that AT did not confer increased reactivity, as the 'head to tail' bisadduct was still produced. Instead, monoalkylation resulted in a distortion 4 to 5 base-pairs in the 5′-direction, at the 5′-TA step junction. Therefore the bisalkylation reaction occurs at the least warped adenine furthest away form the junction site.

As demonstrated above, unlike the 5′-•TAATTA* sequence which yields an interstrand cross-link (11-47%), 'A' tracts instead accumulate more monoalkylated product *e.g.* 5′-TTTTTA*, 5′-TTTAAA* and 5′-TTAAAA* (14-27%), than cross-linked (2-10%) (Seaman and Hurley, 1996). This is a consequence of the intermolecular rearrangement of DNA within the 5′-TAATTA sequence, where the base pairs either open or rotate into a Hoogstein conformation to stabilize the DNA in an unbent conformation. The second arm of bizelesin can then complete the formation of the interstrand cross-link.

The superior pharmacological efficacy and cytotoxic potency of bizelesin, in comparison to the monofunctional alkylators (+)CC-1065 and adozelesin, have made it an important chemotherapeutic agent for clinical trial.

1.3 Minor groove binding with intercalation

The last major family of minor groove binders to be discussed in this mini-review incorporate an intercalating binding moiety into their mode of action. The binding and sequence selectivity of the anthracyclines towards DNA will now be discussed in some detail.

1.3.1 Anthracyclines

The anthracycline family consists of some of the best known and most chemotherapeutically useful antitumour antibotics. Their clinical activity manifests itself *in vivo* by three mechanisms: DNA intercalation and topoisomerase inhibition *via* the minor groove, aerobic and anaerobic redox activation and membrane association (Fisher and Aristoff, 1988). Each of these may also contribute to the cytotoxic side effects which occur upon clinical application of these drugs, aerobic redox cycling is thought to be the greatest culprit of host induced cardiotoxicity. The main structural feature of these intercalators is the large tetrahydrotetracenequinone chromophore which consists of three flat coplanar, six-membered rings (Fig. 26) (Wang *et al.*, 1995). The fourth ring is the minor groove binding moiety. It is a puckered cyclohexene ring which has various sugars attached to it, which are specific to the individual anthracycline. Collectively, these four rings (A-D) compose the aglycone. This chromophore intercalates between adjacent base-pairs during a transient disruption within their stacking arrangement, allowing the ring system to infiltrate the DNA whilst the sugar moiety binds into the minor groove. Other substituents on the 'A' ring are for stabilization of the complex by hydrogen-bonding.

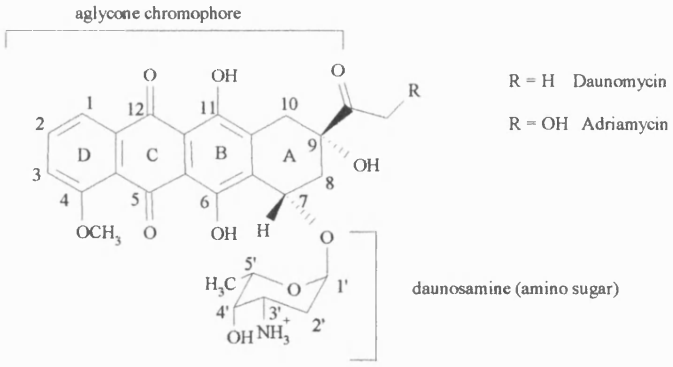


Figure 26 Structures of the major anthracyclines.

Daunomycin (daunorubicin) was first isolated from cultures of *Streptomyces peucetius* (Di Marco *et al.*, 1964). It was found to display weak antibiotic activity against some bacteria but was extremely cytotoxic against tumour cell lines, specifically leukaemia, which has been its greatest clinical application. Adriamycin (doxorubicin) is nearly structurally identical to daunomycin, the extra hydroxyl substituent suggesting that it is a daunomycin metabolite from *Streptomyces peucetius var. caesius* (Arcamone *et al.*, 1969). Compared to daunomycin, adriamycin has a broader spectrum of activity and increased cytotoxic potency, which is reciprocated in its widespread clinical use in cancer chemotherapy against solid tumours and particularly in the treatment of breast cancer. The development of cross-resistance within cells between daunomycin, actinomycin D and Vinca alkaloids (Dano, 1971) is an additional problem to the cardiotoxicity experienced on anthracycline administration. Cross-resistance to adriamycin in daunomycin-resistant L1210 tumour cells suggested that both compounds had a shared mode of action.

The high cytotoxic potency of adriamycin corresponds to the high binding affinity it exhibits for DNA on intercalation, making it one of the anthracyclines most securely fixed to DNA (Capranico *et al.*, 1986). Binding is more stable than that produced by daunomycin, due to the presence of the hydroxyl which is capable of forming a hydrogen-bond *via* a water molecule to the phosphate backbone of DNA (Wang *et al.*, 1987). Footprinting data demonstrated that intercalation occurs between adjacent guanine and cytosine (d(CG) or d(GC)) base pairs preferably flanked by an AT base-pair with the sequence 5′-(A/T)CG favoured over 5′-GCG (Chaires *et al.*, 1987). These consensus sequences are often found in promoter and enhancer regions, where modification of DNA could severely affect transcription. Experiments with daunomycin have shown there is little effect on RNA synthesis. The main autitumour effect is exerted through inhibition of DNA synthesis. This is because anthracylines have the ability to competitively inhibit DNA polymerases attachment to the DNA receptor site, as well as inhibit chromatin compaction (Sen and Crothers, 1986).

Wang *et al.*, (1987) examined the structure of the daunomycin-5′-d(CGTACG) adduct. The anthracycline occupied the CG base pairs, with the aglycone moiety intercalated such that its long axis was approximately perpendicular to that of the DNA helix. The methoxy 'D' ring protruded into the major goove, whilst the cyclohexene 'A' ring projected into the minor groove. By itself, the aglycone is not an anticancer agent, suggesting the importance of the charged amino sugar which seems to stabilize the initial complex, allowing the aglycone to thread between the base-pairs. The configuration of the attachment site where the amino sugar is bonded to the 'A' ring is important for conferring right hand chirality in the molecule, making its conformation complementary to the β-helix. This glycosidic linkage is quite rigid. Unlike adriamycin, daunomycin forms no hydrogen-bonds with the minor groove. However, some stabilizing effects are apparent from the 'A' ring anchor at C9. The acetyl group is within hydrogen-bonding distance of G-N2 and the hydroxyl forms a water-bridged hydrogen-bond to cytosine C=O. These interactions result in a long residence time of the drug and hence high cytotoxic potency. Wang *et al.*, (1995) explained the preference for the 5′-(A/T)CG consensus sequence in terms of a possible clash between the N of the amino sugar and the NH_2 of G, if the guanine was on the 5′-side of the intercalation site. This would result in a less stable adduct. The possibility of this collision was, nevertheless, exploited to form a methylene cross-link by nucleophilic attack of formaldehyde on the daunomycin 5′-d(CGCGCG) adduct. It was hypothesised that this entire complex could be used as an anticancer agent, the DNA acting as a carrier for the drug which would be released slowly *in vivo* on the cleavage of the aminal linkage, perhaps prolonging the cytotoxicity of the drug. Interestingly, administration of the daunomycin-DNA-HCHO complex and free daunomycin, yielded different results relative to the dosage (Wang *et al.*, 1995). The free daunomycin killed the L1210 cells within 3 days. Low concentrations of the complex produced larger cells which subsisted until day 5, whilst high concentrations of daunomycin-DNA-HCHO caused cell shrinkage and cell death within 48 hours, suggesting increased antineoplastic activity. Because DNA is non-immunogenic, this complex could hold the key to a new line of cancer chemotherapy.

Topoisomerase II inhibition is as important as DNA polymerase inhibition, as a mechanism through which anthracyclines can exert their cytotoxic efects. Anthracyclines possibly alter the structure or population of topoisomerase II in a manner which results in the loss of the topoisomerase-mediated protein-DNA linkage. N-Trifluoroacylation of the sugar amino moiety removes the positive charge from the ligand and, therefore, its ability to intercalate, but the antitumour activity remains in the cell by topoisomerase II inhibition (Fisher and Aristoff, 1988). This mode of action is not clearly understood.

The amino sugar of anthracyclines is the moiety of choice to modify, when attempting to vary pharmocokinetic properties or introducing alkylating groups. This is because alteration

of the anthraquinone to reduce its redox potential not only results in the desirable curtailment of cardiotoxic side effects but also loss of cytotoxic potency. The quinone is a functional group in many antitumour agents. Cullinane *et al.*, (1994) tried to elucidate the role of the bioreductive activation process in the formation of adriamycin-DNA adducts. Previous studies *in vitro* demonstrated that anthracycline covalent binding to DNA was *via* C7 of a reductively activated quinone methide. Xanthine oxidase and cellular flavoproteins such as NADH dehydrogenase are enzymatic reducers *in vivo*. However, Cullinane *et al.*, (1994) was able to use a DTT/Fe reducing system in which addition of $1e^-$ produced a semiquinone, another e^- yields the hydroquinone which leads to the quinone methide (tautomeric with C7 carbocation). It is this species which can react with nucleophiles, such as guanine N-2, forming a 7-deoxyaglycone linkage (Fig. 27).

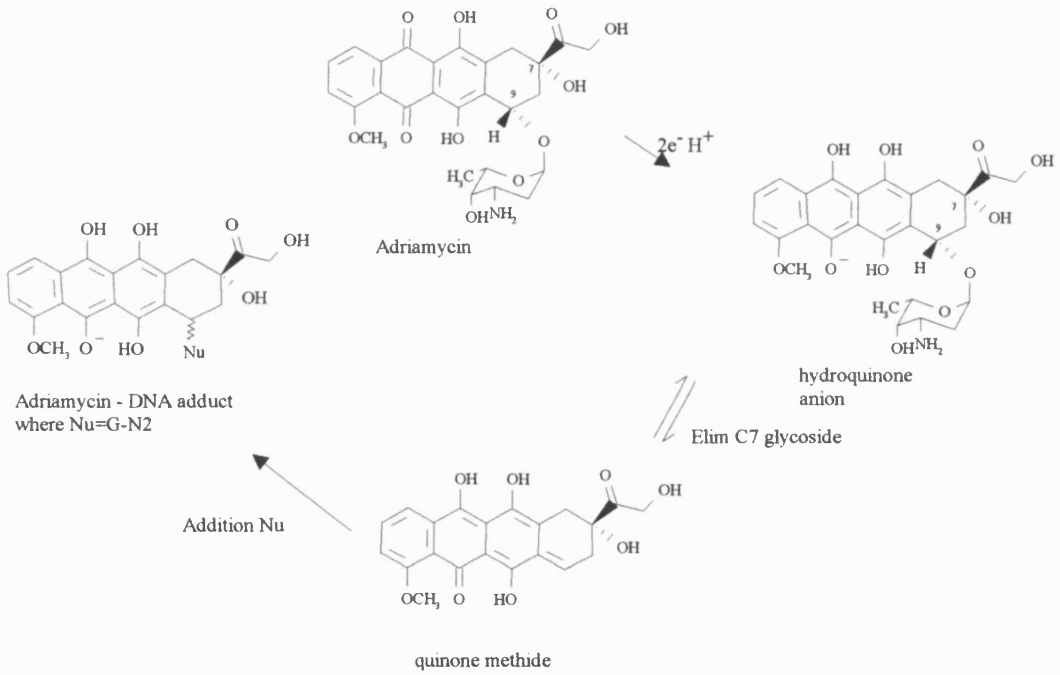


Figure 27 Reductive activation of adriamycin.

The C9 position could be modified extensively, as in other anthracyclines, and the reaction would still proceed to form an adduct. No adduct has yet been isolated to verify this theory, due to their heat and alkali instability. Van Rosmalen *et al.*, (1995) examined the stabilities of this adduct *in vitro* and deduced its half-life to be approximately 40h at 37°C. The covalent attachment of adriamycin to DNA has a half-life of ~5h, although this does not lead to the immediate release of the ligand from the DNA. The four rings of the chromophore remain strongly intercalated, which results in the long half-life of this adduct.

3′-Deamino-3′-(3-cyano-4-morpholinyl)adriamycin (CMA) is a synthetic analogue of adriamycin (Acton *et al.*, 1984). A cyano group is attached to the morpholinyl ring which also

incorporates the sugar amine (Fig. 28). This yields a compound with antitumour activity which far exceeds that of its parent. CMA is between x100→x1000 more cytotoxic than adriamycin against P388 leukaemia in mice and from x100→x500 more potent against HT-29 human colon carcinoma. The fact that these cancers and human ovarian sarcoma are not resistant to CMA suggest that it may have a different mechanism of action to adriamycin. CMA is of significant clinical importance not only because of its extreme cytotoxicity but due to its lack of cardiotoxicity. Begleiter and Johnson (1985) studied the interaction of CMA with HT-29 cells using an alkaline elution assay. DNA cross-linking was observed which increased with drug concentration. Equally cytotoxic concentrations of adriamycin however, produced no cross-links suggesting that this feature was attributable to the cyanomorpholinyl moiety.

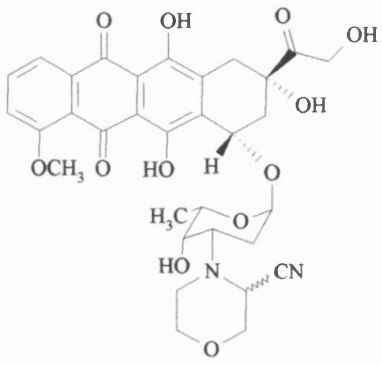


Figure 28 The structure of CMA.

CMA, the 5-imino derivative of CMA (ICMA) and chlorambucil were compared on their reaction with DNA, to try to elucidate the roles of the different functional groups in CMA on binding (Jesson *et al.*, 1989). CMA was more potent and produced a higher level of cross-linking than ICMA, suggesting that the quinone is employed in forming cross-links. Chlorambucil was unable to intercalate into DNA, unlike CMA and ICMA, but chlorambucil was still able to perform cross-linking in the presence of ethidium bromide when CMA was inhibited from so doing. Intercalation is therefore important for CMA activity. This was also confirmed by diminished intercalation on reduction of the quinone, possibly due to the lack of stabilizing hydrogen-bonds which form on complexation with DNA. Westendorf *et al.*, (1989) detected the loss of CN from [^{14}C]N-morpholinoadriamycin during cross-link formation with DNA, as Acton *et al.*, (1984) had predicted yielding the reactive iminium ion, which would react well with nucleophilic centres of G-N2 and G-N7. Other modifications to the N*-position on the morpholinyl ring also blocked cross-linking, leading to the conclusion that an intact morpholino ring was required for adduct formation.

In vitro transcription assays on the CMA-DNA adduct showed transcriptional blockages at 5′-CC and 5′-GG sites (Cullinane and Phillips, 1991; 1992; 1994). Because G-N2

or G-N7 are susceptible to alkylation by other ligands it was assumed that CMA did likewise, the 5′-CC blockage arising from being opposite the binding site. This consensus sequence results in intrastrand cross-links, although some interstrand cross-links were noted at 5′-GC sites *in vivo*. More sensitive assays were required to detect interstrand cross-links *in vitro* from HeLa cells, due to their low population. Intercalation of CMA at these interstrand cross-linking sites was examined by NMR (Odefey *et al.*, 1992), using the sequence 5′-d(CGTACG). CMA inserted itself between the CG base-pairs at either end of the oligomer, with the sugar-morpholine side-chain extending further into the groove than seen with the parent anthracyclines, nearer to an AT base-pair. The stabilizing hydrogen-bond previously seen with adriamycin was also present here, between O9 hydroxyl and the N2 of guanine. Further NMR studies have been performed on this sequence and 5′-d($C^1G^2A^3T^4C^5G^6$) to review the binding of (2S)-2-methoxymorpholinodoxorubicin and morpholinodoxorubicin (Mazzini *et al.*, 1998). NOE data showed intercalation of the aglycone between the terminal CG bases. H1 and H3 of the aglycone 'D' ring showed NOEs with C^1 and $C^{11}H5$ within the major groove. The 11-OH and 6-OH interacted with C^1 and G^{12} respectively. As with adriamycin, the 9-OH is in hydrogen-bonding distance of both G-N2 and G-N3. Meanwhile, the daunosamine ring is situated in the minor groove from the NOE coordinates of A^3/T^3 and G^{12}, depending on whether the adduct formed is with the 'TA' or 'AT' containing sequence. Daunosamine is closer to G^2 and T^3 in the 'TA' sequence, whilst it migrates towards the other strand, *i.e.* C^{11} and T^{10}, in the 'AT' oligomer. The morpholinyl ring showed NOEs to T^4 or A^4 in a similar manner, which could be indicative of a water-mediated H-bond from the hemiacetal proton (Nβ) to the deoxyribose O4′, enhancing the residence time of this drug and consequently the potency.

Ettore *et al.*, (1998) studied the X-ray crystal of CMA and the DNA hexamer 5′-d(CGATCG). As reported by Odefey *et al.*, (1992) and Mazzini *et al.*, (1998), CMA intercalated between the terminal 5′-CG base-pairs, forming a hydrogen-bond to guanine. However, the structure of the morpholino moiety was observed not only to have lost the cyanide group but also to have opened its ring to form N-(2-hydroxyethyl) adriamycin. No alkylation took place perhaps because the morpholino group was adjacent to an AT base pair rather than one containing a guanine NH_2.

Both inter- and intrastrand cross-links of DNA are heat labile (Cullinane and Phillips, 1993), which correlates well with the idea that the cross-link is composed of an aminal linkage N-C-N, from 3-position of the morpholino ring and the N2 of guanine. The N2 of guanine is implicated over N7 because λ exonuclease detects only minor groove blockages, unless these is a large deformation of the helix originating from the major goove which is unlikely. The true mechanism of cyanomorpholinoadriamycin cross-linking has still to be elucidated.

Nogalamycin was first isolated from *Streptomyces nogalator* by Bhugund and Dietz (1965). Some antibiotic activity is exhibited against gram positive bacteria, however, it is its potency against a number of tumours, including L1210 leukaemia cell lines, which is most significant. Renal and pulmonary toxicity precluded its clinical development. Nogalamycin is a large anthracycline with substituents protruding from both ends of the aglycone chromophore. An uncharged nogalose sugar is attached to the 'A' ring, whilst a positively charged bicyclic aminoglucose sugar is joined to the 'D' ring (Fig. 29). The mode of action of nogalamycin is through selective inhibition of DNA-directed RNA synthesis (Li *et al.*, 1979). Nogalose is thought to be closely connected with an enhanced inhibition of the transcription process or RNA synthesis, which leads to retardation of cell growth followed by cell death. Cytotoxicity occurs in all phases of the cell cycle but is most pronounced in 'S' phase.

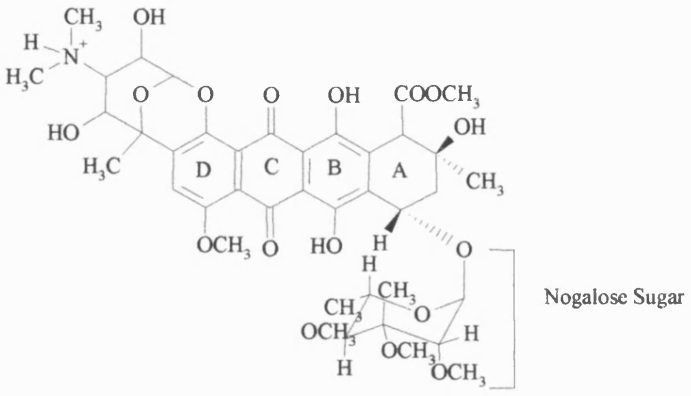


Figure 29 Structure of nogalamycin

An NMR study of the nogalamycin-5′-d($G^1C^2A^3T^4G^5C^6$)$_2$, complex indicated that two ligands bound per oligomer between the d(CA) sites. The nogalose sugar lays in the minor groove while the charged bicyclic aminoglucose sugar is situated in the major groove, both pointing towards the terminals of the duplex (Searle *et al.*, 1988). Rearrangement of the base sequence of DNA to 5′-d($C^1G^2T^3A^4C^5G^6$) resulted in the intercalation occuring at d(CG) instead. The sugar groups now face the centre, which is likely to be energetically more favourable than the aglycone arms flailing freely in solution (Robinson *et al.*, 1990). Theoretically, because of this drugs preference for 5′-NpG or 5′-CpN steps, with the aglycone intercalating on the 3′-side of C or the 5′-side of G and the sugars facing GC, nogalamycin should be able to bind to these sequences in either direction. However, oligomer size constraints are the controlling factor. Alternate pyrimidine-purine (3′-5′) steps are thus required for the intercalation of the aglycone of nogalamycin. A distortion of the DNA helix was noted at C^5-G^6 (and G^{11}-C^{12}) from an enhancement of NOEs between them. This suggests that C^5 buckles towards G^6 on nogalamycin binding, encasing the aglycone to form a wedge shaped binding site (Robinson *et al.*, 1994). Gao *et al.*, (1990) noted from X-ray diffraction analyses that bound nogalamycin spanned three base pairs, with the seemingly planar aglycone flexing so

as to increase the quality of the complementary fit with DNA. This maximizes hydrogen-bonding of O7 and the carbonyl (of C10) with the minor groove. There are also van der Waals forces available in the major groove for additional stabilization of the complex.

One of the most surprising facts of the mode of antitumour activity of nogalamycin is how it manages to intercalate into DNA with such large sugar moieties projecting from both ends of the aglycone, which result in a dumb-bell shaped molecule. The diameter of methyl ester to nogalose is 12Å, and the width of the aminoglucose sugar is 6Å, while the distance between adjacent base-pairs is only 3-4Å (Smith *et al.*, 1996). This makes it impossible for either end to thread through DNA without some form of base-pair opening. Although DNA could flex or unwind, the most likely event is transient melting of the helix, where an arm of nogalamycin can slip through and the helix effectively then zips itself up around the aglycone. A mechanism such as this is compatible with the comparatively slow rate of binding compared with other anthracyclines. However, these sugars are also efficient molecular anchors, increasing the residence time of the drug and consequently the cytotoxic potency. Nogalamycin is one of the few naturally occuring antibiotics which exerts its action by binding in both the minor and the major groove simultaneously.

Successful analogues of nogalamycin have been synthesized, such as menogaril, which is active against a wide spectrum of tumours and has proceeded to Phase II clinical trials. Menogaril lacks the acetyl and nogalose sugar functionalities, which were perhaps responsible for the toxic side effects seen with nogalamycin. This ligand intercalates with DNA from the direction of the major groove.

1.4 Summary

It is clear, from this review, that the binding of ligands to DNA is an exceedingly complex process. Binding strategies not only display enormous diversity between molecular families but also within them, depending on the presence of different functional groups. There are, of course, many other antitumour antibiotics which incorporate themselves into the minor groove such as: the aureolic acids (*e.g.* chromomycin and mithramycin), the quinoxalines (*e.g.* triostin A and echinomycin), the duocarmycins, the microgonotropens, acridine base compounds (*e.g.* AAC and DACA), short peptide chromosomal fragments, ecteinascidins, mitomycins, enediynes (*e.g.* calcheamicin and esperamicin), glycopeptides (*e.g.* bleomycin) and, lastly, pluramycin antitumour antibiotics such as pluramycin, hedamycin and altromycin (Fig. 30). The characteristics of their binding have not been described here to limit the length of this review.

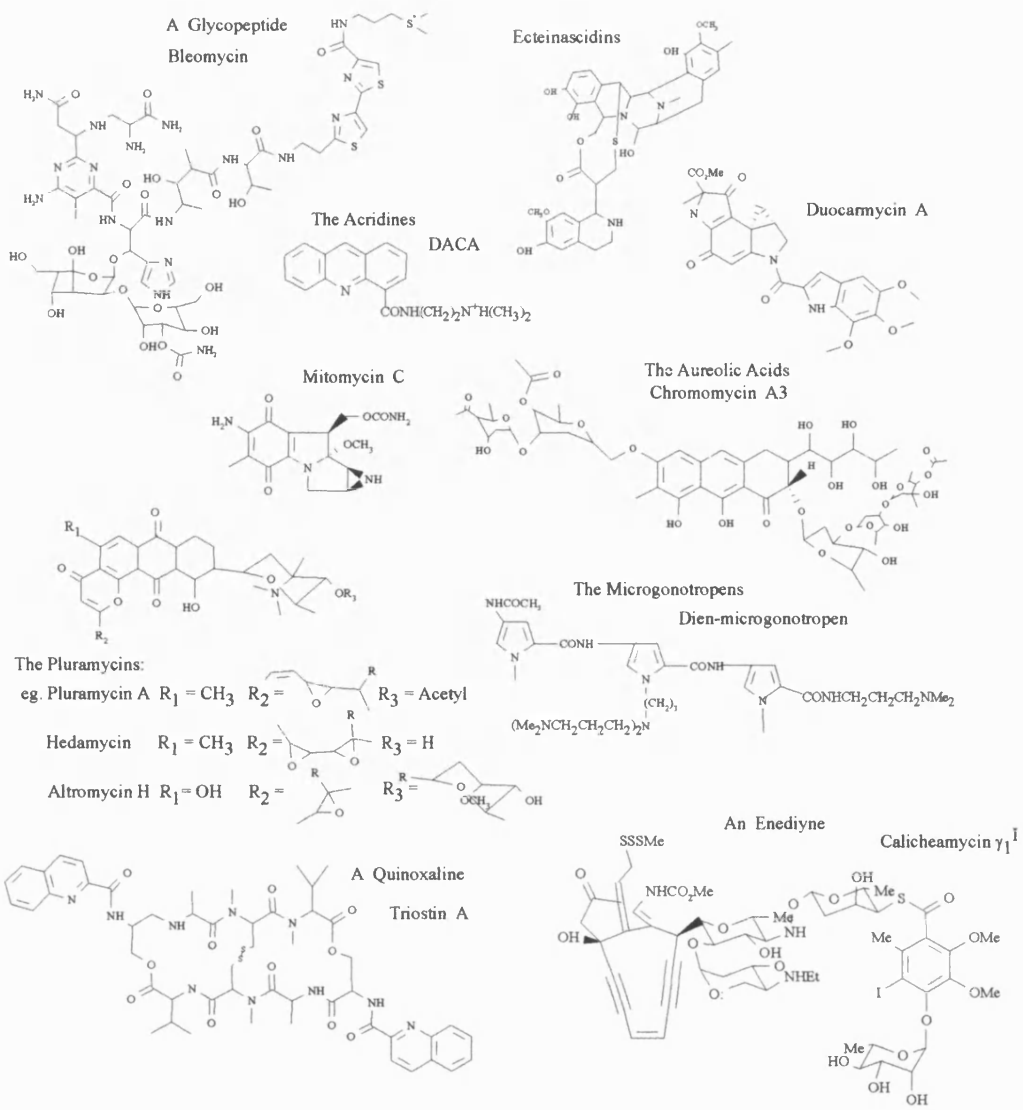


Figure 30 Structures of some common minor groove binders.

In summary, ligand binding to the DNA minor groove is controlled by a host of contributing factors:

- Minor groove geometry, narrow 'A' tracts or wider GC rich regions.
- Hydrogen-bonding, electrostatic or hydrophobic interactions, van der Waals forces, bifurcating bonds.
- Presence of G-N2 adds a steric obstacle to some ligands binding.
- β-helix right hand twist is complementary to that of many naturally occuring binding agents.
- Does the ligand require intercalation of a chromophore to form an anchor, π-π stacking of aromatic drug-DNA rings?
- Recognition of an unusual DNA feature; increased helical twist, melted DNA, bent DNA, 5′-TA junction site.
- Presence of nucleophiles or electrophiles for covalent bond formation.

- Flanking base pairs of the binding site.

There appears to be no simple general recognition code to explain the binding of small molecules to the minor groove of DNA, which is why DNA-targetted drug design is exceptionally problematic.

This introduction has only considered ligands which interact with DNA *via* the minor groove. There are however, many drugs which target the major groove of DNA. Nitrogen mustards *e.g.* cyclophosphamide and chlorambucil, form a cross-link by alkylating the N7 of guanine in adjacent base pairs. Nitrosoureas *e.g.* carmustine, and busulphan also alkylate the major groove of DNA. Cisplatin is perhaps one of the most well-known alkylating agents which produces intrastrand cross-links between N7and O6 of adjacent guanine molecules within the major groove, resulting in the breakage of the hydrogen-bonds between GC bases and thus denaturation of the local DNA chain. Not only small ligands interact within the major groove, oligonucleotides of up to 25 base-pairs in length can insert themselves into this site, adhering through hydrogen-bonding. The resulting triple helix structure is particularly interesting because it leads to the possibility of using oligonucleotides, to target specific gene sequences in a therapeutic manner.

1.5 Aim of this project

The aim of this project was to investigate the binding of various antitumour agents to DNA at the molecular level. As reviewed in the previous sections of this chapter, a large volume of information is already available on the affinity, efficacy and sequence-selectivity of various families of drugs towards their intercellular target DNA. However, the picture is far from complete, with only a few drugs from each family having their mode of binding examined at the molecular level by high field NMR or X-ray crystallography. These studies often only investigate the information of a drug-DNA complex at a few DNA target sequences. This makes it difficult to establish a trend of requirements for the sequence-selective binding mode, either from the perspective of the drug by itself or more usefully the entire drug family.

Studies using techniques such as gel electrophoresis have already established an answer to the question: what DNA consensus sequence does a particular drug prefer? This includes the proposed site of alkylation or intercalation, plus the 5′- or 3′- flanking sequence to this site, for example; adozelesin prefers 5′-(A/T)(A/T)A*(C/G)G or p(A) (Weiland and Dooley, 1991). The question this project tries to answer is: why does the drug, *e.g.* adozelesin, prefer this sequence? Only a close investigation of the molecular structure of the adozelesin-DNA

adduct by high-field NMR and restrained molecular modelling, can provide an insight into why adozelesin favours these sequences over all other combinations of nucleotides. It was hoped that scrutinization of the refined molecular model of this drug-DNA adduct, which represented the NMR data, would reveal a number of key points relevant to the similarities and differences between adozelesin and other members of the cyclopropapyrroloindole family, or at a more general level compared to all other minor groove binders. This description of the binding mode would constitute; the alkylation site, the flanking sequence and its length, the possibility of selective noncovalent interactions or stabilizing hydrogen-bonds within the oligomer sequence, or if the DNA has some unusual feature which is conducive to drug binding *i.e.* the conformation could be straight or bent, wide or narrow.

Characterization of drug-DNA interactions was only possible after a long sequential preparatory procedure, which included:

1) The synthesis and purification of a DNA oligomer, using HPLC and hydroxylapatite chromatography. This was followed by collection of a 1D ^{1}H NMR spectrum to check if the correct oligomer was present in a pure form. If the oligomer was non-self-complementary, a complementary strand was prepared in the same manner and they were then annealed to yield DNA duplex. This DNA sample was then analysed by two-dimensional NMR.
2) The production of a single pure drug-DNA adduct, from the reaction of the ligand with a previously synthesized and purified DNA oligomer. Purification of the drug-DNA adduct was achieved using C^{18} Sep Paks and hydroxylapatite chromatography. This was followed by 1D ^{1}H NMR analysis to verify the drug-DNA reaction had proceeded to completion, yielding a single drug-DNA adduct.
3) Collection and analysis of two-dimensional ^{1}H NMR data, implementing mainly DQF-COSY and NOESY experiments, with some input from ROESY and TOCSY data. Once the assignment of all non-exchangeable drug and DNA protons within the adduct had been completed, distance-range constraints were generated and listed within a peak-pick spreadsheet.
4) A computer molecular model of the hypothesized structure of the drug-DNA adduct was built, by docking a minimized adozelesin molecule into the minor groove of canonical DNA. The distance-range constrants were applied to the structure, which was then refined by molecular mechanics and dynamics calculations

.

A number of antitumour antibiotics were used, owing to the different degrees of success experienced in the second stage of this project. The ligands for which the structures of their corresponding DNA adducts were being scrutinized were; cyanomorpholinoadriamycin,

sibiromycin, DSB-120 and adozelesin. Information on the precise conformation of these adducts could provide improved information for more fruitful drug design.

1.6 Why choose NMR techniques for structure analysis?

Over the last fifteen years, especially, huge strides have been taken in the development of NMR technology which is powerful enough to allow the elucidation of three-dimensional biomolecular structures in solution. These technological advances include: the availability of superconducting high-field NMR spectrometers with their modern computer systems and field strengths of 9 to 14 Tesla (400 and 600MHz resonance frequency for the proton nucleus) and the invention of multi-dimensional NMR which utilizes various pulse sequences to collect data on nuclear spin connectivities. The introduction of increasingly high fields corresponds with enhanced sensitivity, which is perfect for the analysis of complex biological macromolecules, that are typically only available at low concentration. Multi-dimensional NMR data provides conformational information *via* through-bond and through-space connectivities. NMR has a wide variety of applications not only in refining molecular conformations but also in characterizing binding and dissociation rates, plus intermolecular interactions between a ligand and a receptor. Measurement of 2D NOE cross-peak intensities was performed in this investigation to generate internuclear distance constraints. These were subsequently used in a molecular modelling study to elucidate the 3D conformational structure of the DNA adduct in question.

Much discussion has occurred about the relevance of X-ray crystal structures, such as those of oligonucleotides, to molecular conformation in the solution state (Westhof *et al.*, 1995). Crystal packing interactions between the ends of DNA duplexes are thought to produce perturbations in the helix terminal structure, so that only the central 8 base-pairs of a 12-mer are structurally accurate. However, this could also produce an environment more like that experienced naturally in the long DNA strands within a chromosome. In solution, these bases have a propensity to fray, as can be detected by NMR. Generally, there is good correlation between both sets of data although some differences are apparent *e.g.* hydrogen-bonding. The premier advantage of NMR structural analysis is that, the sample is in solution state, where parameters such as pH, temperature and ionic strength can be varied. This is especially important for examination of biological molecules such as DNA, protein, RNA and carbohydrates, which for true structural definition require conditions mimicking those they are subjected to *in vivo*.

1.7 The Nuclear Overhauser effect

The nuclear Overhauser effect (NOE) can be defined as the transfer of magnetization between two spins A and B, which are in close spatial proximity (<5Å) by way of through-space cross-relaxation. The irradiation of one nucleus at its resonance frequency leads to weakening or strengthening of the signal due to the second nucleus. This interaction is known as nuclear Overhauser enhancement and corresponds to the relaxation of the observed nucleus by the irradiated nucleus, as they tumble about each other at a frequency close to that associated with relaxation. There is no requirement for scalar coupling J_{AB} between the nuclei. The nuclear Overhauser effect therefore, produces a further increase in the signal to noise ratio.

The nuclear Overhauser effect can be easily illustrated using the two spin system of an AB spectrum *e.g.* C-H where A and B are relaxing with one another resulting in four populated energy levels (Abraham *et al.*, 1988; Evans, 1995; Williams and Flemming, 1987) (Fig. 31a). Only spin $^1/_2$ nuclei are used, where both A and B can exist in either $+^1/_2$ or $-^1/_2$ orientations *i.e.* indicating the alignment of the nucleus enforcing or opposing the applied field. W denotes the probability of a single quantum transition, the subscript *e.g.* W_1 gives the change in quantum number, and A or B refers to the nucleus. There are six possible transitions in Figure 31a, of which W_{1A} and W_{1B} are allowed and give rise to the observed NMR lines, whilst zero (W_0) and double (W_2) quantum transitions are 'forbidden' in the selection rules so cannot be seen. The probability of the W_{1A} transition is proportional to the difference in population between the connected energy levels, *i.e.*

$$E_4 - E_3 = (^N/_4 + n) - (^N/_4) = n \text{ or } E_2 - E_1 = (^N/_4) - (^N/_4 - n) = n$$

This also applies to W_{1B}.

Irradiation at the resonance frequency (δ_B) of nucleus B (Fig. 31b), during a time period τ causes saturation of this resonance. The irradiation is followed by a nonselective irradiation pulse and acquisition of the free induction decay (FID). At saturation the energy levels connected by the W_{1B} transitions have their populations equalized *i.e.*

$$E3 = E1 = {}^N/_4 - n \text{ and } E2 = E4 = {}^N/_4 + n$$

where the population levels of E1 and E2 grow at the expense of E3 and E4. At this point the intensity of the A signal remains unaffected as it is a result of the transitions from E1 to E2 and E3 to E4. The intensity of A is controlled by the difference in populations between E1 and E3 and of E2 and E4 which remain the same. Relaxation can occur along the reverse of the transition pathways, but more effective is the emission of radiation which arises *via* zero (W_o) and double quantum transitions (W_2). W_2 occurs between the most well separated energy levels and is favoured by the high frequency tumbling of molecules with a low molecular weight. The

population of E4 increases at the expense of E1 hence, the populations of energy levels E1 and E3 are less than that of E2 and E4, resulting in enhancement of the signal from the A nucleus.

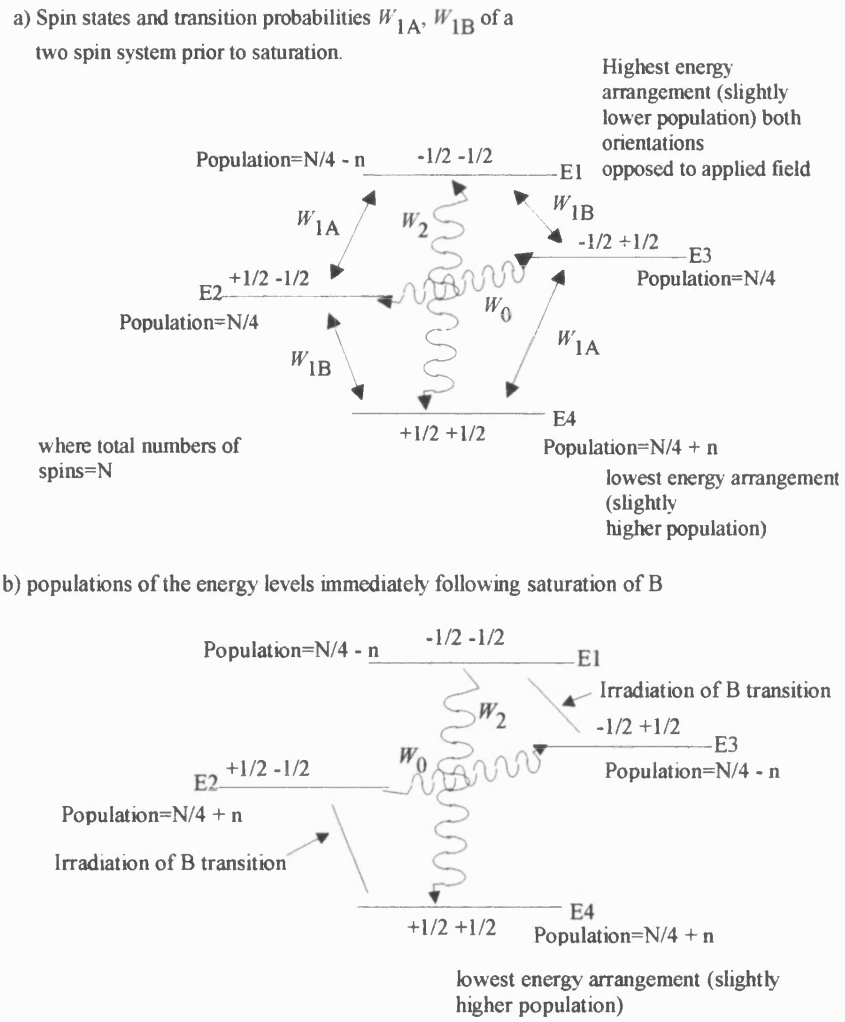


Figure 31 Energy level populations before and after irradiation in an AB system.

Zero quantum transitions (W_o) are common in larger molecules with slower tumbling rates. The population of E2 is increased at the expense of E3 (an energy level of similar energy) leading to an increase in the populations of E2 and E1 relative to E3 and E4. A reduction in the intensity of the A nucleus signal results from radiation emission *via* W_o. This brings the molecule back to the ground state, maintaining the Boltzman distribution between the $-^1/_2$ $-^1/_2$ and $+^1/_2$ $+^1/_2$ spin states. It is these double and zero quantum transitions after irradiation of nucleus B, which result in a change in the population distribution of spins in nucleus A, hence giving rise to the observable NOE. The difference in the populations of W_{1A} after these transitions, is denoted as;

$$(E_2 - E_1) = (^N/_4 + n) - (^N/_4 - n) = 2n \text{ or } (E_4 - E_3) = (^N/_4 + n) - (^N/_4 - n) = 2n$$

leading to an increase in NOE intensity by a factor of 2.

Nuclear magnetic dipole-dipole coupling between nuclei which results in an NOE, is a consequence of the local field at one nucleus due to the presence of a field at another, as molecules move around in solution. Homonuclear decoupling is described using the equation:

$$\frac{i'}{i} = 1{\cdot}0 + \frac{\gamma_B}{\gamma_A}$$

← decoupled B nucleus
← observed A nucleus

Where:
i' = intensity of the A transitions in an AB spectrum when B is decoupled
i = intensity without coupling

When $\gamma_A = \gamma_B$, the maximum observed NOE is 0·5 (50% enchancement). In the case of heteronuclear decoupling *e.g.*

$^{13}C - H$: $\frac{i'}{i} = 1{\cdot}0 + 0{\cdot}5$ $\frac{\gamma^1H}{\gamma^{13}C}$ ←~ ¼ of the ^1H magnetogyric ratio

a three-fold enhancement maybe possible in the signal to noise ratio. Molecules moving in solution encounter different dipolar interactions, depending on molecular orientation and internuclear distance which may stimulate cross relaxation and hence an NOE.

NOE intensity is proportional to the distance (r) between the irradiated nucleus B and the observed nucleus A, which is a function of correlation time τ_c (takes into account the motional averaging process on the observed NOE). This gives rise to the following equation:

$$NOE \alpha \frac{1}{r^6} \cdot f(\tau_c)$$

The NOE intensity is not always a true reflection of distance. Problems with low sensitivity for observation (low signal-to-noise ratio), make it more difficult to calculate integrals. Large molecules encounter further problems because they do not rotate as fast in solution (τ_c (rotational correlation time) » W_0^{-1}(Larmor frequency)), so spin diffusion pathways may become as prominent as cross-relaxation changing the intensities of the peaks (Fig. 32).

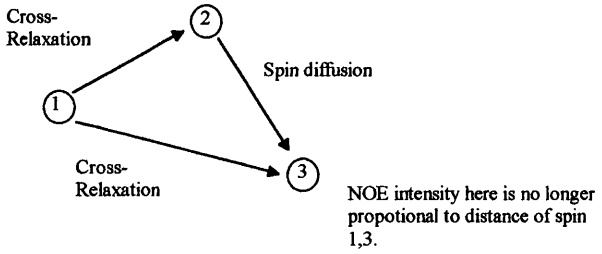


Figure 32 Three spin system exhibiting cross-relaxation and spin diffusion (Wüthrich, 1986)

The magnetization transfer between the nuclei ② and ③ complicates the interpretation of distances from NOE intensities. Increasing the mixing time makes longer range NOEs more observable but spin diffusion is then more prominent in short-range contacts.

1.8 Two-dimensional NMR experiments

All two-dimensional experiments incorporate an incremental delay, known as the evolution time (t_1), into their pulse sequences (Frenkiel, 1993). This incremental delay has a range of values from 0 to t_{1max}, and occurs between an initial $\pi/2$ pulse and an identical second one. At each t_1, an FID (free induction decay) is subsequently recorded in the detection period, which is defined as t_2. The complete data set is then Fourier transformed with respect to t_1 and t_2 to the frequency domains F_1 and F_2, which are linked to the proton chemical shifts in the 2D spectrum. Each dimension exhibits absorption mode lineshape. COSY, TOCSY and NOESY experiments are the most important tools for macromolecular structure characterization.

The double quantum filtered correlation spectroscopy (DQF-COSY) experimental pulse sequence is shown below (Fig. 33):

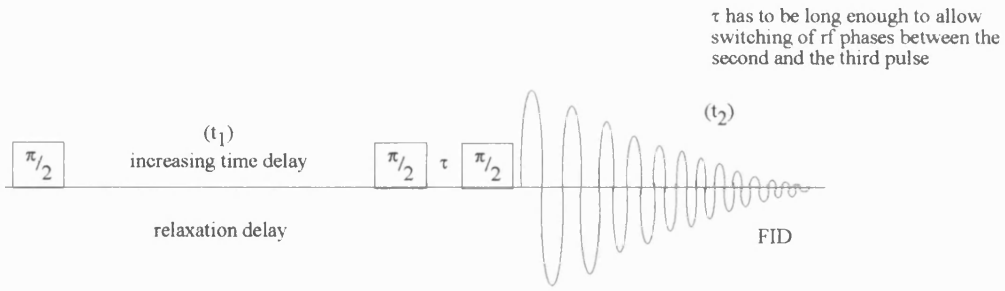


Figure 33 DQF-COSY pulse sequence.

The COSY experiment (minus the second 90° pulse in the above diagram) provides information on protons in a molecule which are scalar coupled. The components of the cross-peak in the spectrum are antiphase which means that, unless the interaction between protons is very strong, the cross-peak will be difficult to detect in the spectrum. Cross-peaks in the spectrum correspond to the nuclei which are coupled to each other and they also contain coupling constant information.

The DQF-COSY experiment uses double quantum filters to improve phase-sensitive COSY spectra. In COSY spectra, diagonal peaks are in dispersion when the phase is adjusted to provide absorption lines of cross-peaks, this leads to the edges of the diagonal dispersion lines tailing off to the baseline slowly. Any peaks such as methyls, which reside close to the

diagonal could therefore be obscured by the wide diagonal fringes (Wüthrich, 1986). Double quantum filtering uses a four-phase cycle to provide axial peak suppression, resulting in antiphase absorption modes of diagonal and cross-peaks. The fine structure of the scalar coupling in the macromolecule is then revealed fully by the cross-peaks.

The fundamental basis, *i.e.* the nuclear Overhauser effect, of the NOESY experiment has already been explained, as has its purpose to identify pairs of protons which are close enough in space to undergo cross-relaxation. The pulse sequence is as follows (Fig. 34):

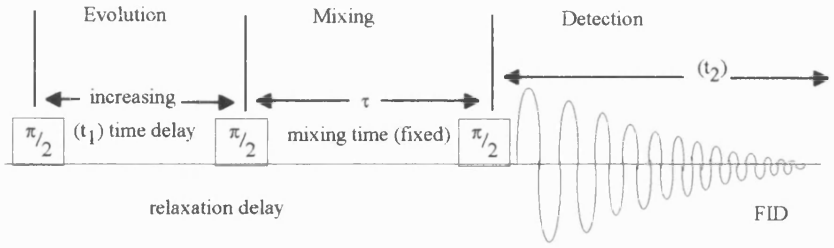


Figure 34 The NOESY pulse sequence

The length of the delay (τ) is chosen according to the rate of the transfer process, NOEs need time to build up, which in turn is dependent on the size and shape of the molecule. Small molecules have long t_1's as they move quickly around solution, taking a long time to relax, so τ is long (*e.g.* 1→5s) to allow NOE build up. The reverse occurs for large molecules which tumble slowly in solution so τ = ~50 → 300ms. After Fourier transformation of the FID, a spectrum is produced with cross-peaks the same sign as the diagonal, which can be +ve or -ve depending on whether relaxation favours W_2 or W_0 routes. No mixing time is required in the COSY experiment because scalar coupling information is already available.

Extensions of the COSY and NOESY techniques are the so called rotating frame experiments; TOCSY (HOHAHA) and ROESY (CAMELSPIN). Cross-peak information as with the previous experiments, is in the form of scalar couplings (TOCSY), and through space interactions (ROESY). However, data build up is now dependent on how these nuclear spins behave in a radio-frequency field.

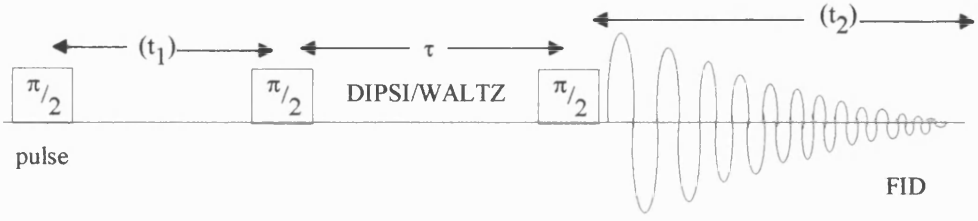


Figure 35 The TOCSY pulse sequence.

TOCSY relies on the exchange of magnetization between coupled spins, by applying a pulse train which eliminates the effect of the chemical shift but retains scalar coupling (Frenkiel, 1993). This is known as 'isotropic mixing' and is illustrated in the above diagram (Fig. 35). DIPSI and WALTZ are the isotropic mixing sequences, that allow continuous irradiation of the spins. In an AB system the transfer of magnetization is optimal at $\frac{1}{2}J_{AB}$, but in macromolecules where there is a large variety of spin systems leading to multistep transfer, there is no general maximum mixing time, so it is approximated instead to 25-50msec for large couplings (single step transfer) and 50-100msec for the contrary. The TOCSY experiment is more sensitive than COSY.

The most significant feature of the ROESY spectrum is that cross-peaks arising from spin diffusion are weaker but are also easier to identify, which is useful when elucidating the structure of macromolecules. Cross-peaks are opposite in phase to those constituting the diagonal. The pulse sequence for this experiment is very simple (Fig. 36):

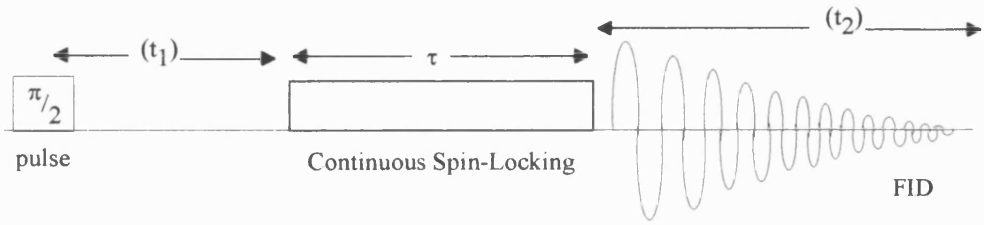


Figure 36 The ROESY pulse sequence.

During the mixing time (τ) a low level radio frequency field (*e.g.* 1-2 KHz) is applied to the nuclear spins, selecting the component of magnetization which is aligned along the irradiation axis, which is otherwise known as spin-locking. The applied radio frequency field is much weaker than the initial pulse. The cross-peaks evolve during the mixing time, the length of which is chosen relative to the rate of cross-relaxation.

1.9 Structure refinement by molecular modelling

Molecular modelling provides a three-dimensional graphical interpretation of the biological macromolecule, on which experimental NMR data has been collected. Computer simulations allow an insight into molecular motion, thermodynamic properties, solvated systems and how ligands and receptors interact, such as a drug with DNA. Knowledge of all these factors, especially within families of drugs, can lead to more successful rational drug design.

The first step to visualizing the macromolecular structure described by the NMR data, is generation of distance constraints using a distance geometry algorithm. These restraints are responsible for 'pulling' and 'pushing' atoms of a starting structure into position during molecular mechanics and dynamics energy refinement. Distance geometry is quite an unbiased method of determining molecular conformation, because constraint application to a variety of starting structures should yield the same result.

Energy minimization is employed to find the local potential energy minima for the macromolecule in question (Evans, 1995). A molecule is in effect a collection of atoms held together by harmonic and elastic forces, which can be described by potential energy functions of structural features such as bond length, bond angle, non-bonded interactions *etc.* (Boyd and Lipkowitz, 1982). These functions are collectively known as the force-field, which is how the potential energy surface for all internal degrees of freedom in a molecule is described mathematically:

$$E = E_s + E_b + E_\omega + E_{nb} + e.g. \text{ constraints}$$

E = the difference in energy between the real molecule and the molecule when all forces are at their ideal values.

E_s = energy of bond stretch/compression from the ideal value

E_b = energy of bond angle bend from the ideal value

E_ω = energy of torsional twist from ideal value along bonds

E_{nb} = energy of non-bonded interactions

There are three standard energy minimization algorithms: steepest descent, conjugate gradient (or Powell) and Newton-Raphson. Steepest descent gradients are especially useful for systems known to have deviated far from the lowest energy state, so they can cope with a rapid reduction in energy *via* large energy steps. This method is often used in conjunction with conjugate gradients, which have much gentler energy slopes, but have the overriding advantage of being able to converge at an energy minima, whilst the former procedure can only oscillate around it. The potential energy surface in all methods is harmonic, so adjustments of atomic positions in Cartesian space leading to a reduction in energy, result in further descent into the well towards the convergence minima.

Steepest descent minimization is defined by the following equation:

$$X_K = X_{K-1} + \lambda_K S_K$$

$3N$ + (cartesian) dimensional vector

where: K = the Kth iteration
λ = eigen value
S = unit length

The molecular co-ordinates are given by the 3N-dimensional vector. Each iteration the molecule is subjected to, involves three steps:

1) Choice of directional descent (using the vector)
2) The size of the descent (controlled by λ_K)
3) The descent itself as described in the above equation

The direction of displacement (S) is straight downhill, parallel to the negative energy gradient, as shown by route A (bold line) when the step size is infinitesimally small (Fig. 37).

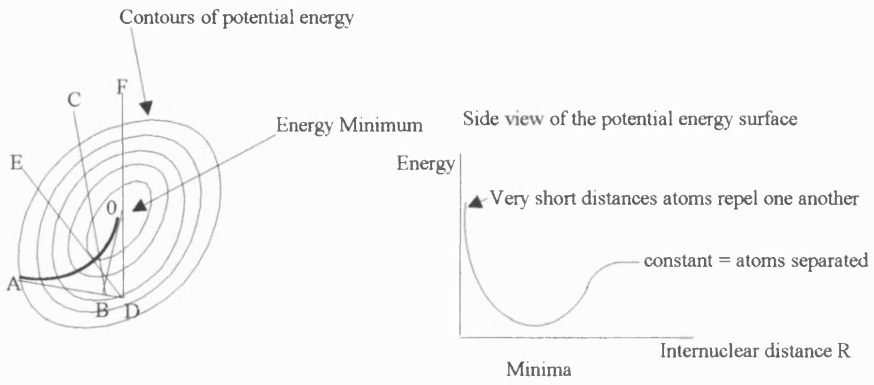


Figure 37 Diagram to show the descent *via* steepest descent (A) and conjugate gradient minimization

The conjugate gradient can also be easily illustrated using this Figure 37. From starting point A, minimization progresses along the negative gradient towards B and D. If it stops at B, the most negative point in this direction, the second step of the conjugate gradient is a combination of the gradient and the previous search direction, as defined by:

$$S_K = -g_K + b_K S_{K-1}$$

Where b_K = weighting factor of the squares of the current and previous gradient
-g = negative gradient

This means the minimization will proceed towards the convergence minima during 'n' iterations. If the first step had arrived straight to D, the second step would be towards F, which would still be closer to the minima than the point arrived at by steepest descent. Steepest descent continues its search in the direction of its previous greatest energy step *i.e.* B→C or D→E for larger step sizes. Hence, conjugate gradient minimization is more intelligent in finding energy minima than steepest descent. However, this latter method speeds up the initial minimization process greatly and the results can then be polished by the former.

Molecular dynamics is another classical 'potential energy'-based technique, used in computer simulation and modelling of macromolecules. This method simulates the vibratory motion of the molecule based on its physical properties, yielding what should be a realistic result. Using the NMR restrained structure which has been minimized, molecular dynamics follows the history from this initial arrangement, by calculating the trajectories of all the atoms

under the influence of intermolecular potentials. Unlike minimization, molecular conformations do not get stuck in local energy minima. So after a period of time ranging from ps to seconds, large areas of conformational space will have been sampled to provide a selection of low energy structures. Higher temperatures mean more conformations are energetically accessible. These structure 'snapshots', are generated from tens of thousands of iterations of calculations which utilize Newtons laws of motion, to predict the location of an atom after a short time period (*e.g.* 10^{-15}s equivalent to a high frequency bond stretch). Large time steps would allow undesirable large movements of atoms. This iterative procedure relieves local stresses corresponding to bond length distortions or non-bonded atom overlaps.

The molecular dynamics simulation tries to solve the equation mathematically (Evans, 1995);

$F_i = m_i a_i$

F_i = force on atom *i*
m_i = mass of atom *i*
a_i = acceleration of atom *i*

The force on atom *i* can be expressed in a differential equation, as a derivative of potential energy V with respect to co-ordinates r_i. For a known mass this can be solved for various positions in time (t_i);

$$\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt_i^2}$$

Because the potential energy is a complex function of the co-ordinates of all atoms in the system, this equation can only be solved generally. The total potential energy V_{total} is defined as:

$$V_{total} = V_{bond} + V_{angle} + V_{dihedr} + V_{vdW} + V_{coulomb} + V_{NMR}$$

Where the first three parameters V_{bond}, V_{angle} and V_{dihedr} restrict bond lengths, angles and dihedral angles to their equilibrium values. V_{vdW} and $V_{coulomb}$ describe van der Waals and electrostatic interactions within the molecule. While V_{NMR} is the potential energy pulling atoms into their correct spatial positions relative to each other, as defined by the NMR generated restraints. Where NOE distances have been sorted into collective 'bins' *e.g.* strong, medium or weak, to produce distance-range constraints, the potential energy function V_{NOE} can be described as;

$$V_{NOE} = \begin{cases} K_{NOE}(r_{ij} - r_{ij}^{u})^2 & \text{if } r_{ij} > r_{ij}^{u} \\ 0 & \text{if } r_{ij} \le r_{ij} \le r_{ij}^{u} \\ K_{NOE}(r_{ij} - r_{ij}^{l})^2 & \text{if } r_{ij} < r_{ij}^{l} \end{cases}$$

where: r_{ij} = calculated interproton distance
r_{ij}^{u} = upper boundary limit
r_{ij}^{l} = lower boundary limit of target distance

K_{NOE} is the force constant which has to be high enough to make sure that the distance-range constraints between atoms within the molecule, are complied with, in the simulation, otherwise an energy penalty is incurred. A value too great could lead to unreasonable demands on molecular conformation, yielding high energies and the molecule 'blowing up'. Convergence at a potential energy minima (there maybe more than one for a large macromolecule, especially in a long simulation where large structural changes may only occur in motions of ps or seconds timescale), demands that NMR restraint violations are at a minimum value.

Molecular dynamics is computationally very expensive in terms of time, which increases linearly with the number of atoms in the system. However, the resulting macromolecular structures are generally in good agreement with the experimental data.

1.10 The DNA-drug complexes discussed in this thesis

Two adozelesin-DNA adducts are discussed in detail in this thesis (chapters 3 and 4). Adozelesin binding to the self-complementary DNA duplex 5′-d(CGTAAGCGCTTA*CG)$_2$, is the first adduct which is characterized in chapter 3. The second drug-DNA adduct investigated was, adozelesin modification of the non-palindromic 'A' tract sequence 5′-d(CGAAAAA*CGG)·5′-d(CCGTTTTTCG). The reaction of CMA, sibiromycin and DSB-120 with various DNA sequences, were also studied and the results are reviewed in chapter 5.

CHAPTER 2: EXPERIMENTAL PROCEDURE

2.1 Laboratory reagents – Sources

β-Cyanoethyl phosphoramidites (deoxy: adenine, thymine, guanine, cytosine and inosine, *i.e.* dA, dT, dG, dC and dI) – Cruachem

Acetonitrile – Cruachem

Iodine (oxidising agent) – Cruachem

Trichloroacetic acid (TCA), (deblock) – Cruachem

Acetic anhydride (Cap A) – Cruachem

4-(Dimethylamino) pyridine (DMAP), (Cap B) – Cruachem

Tetrazole (Activator) – Cruachem

Aqueous ammonia (NH_4OH) – Sigma 29·5% NH_3

Ammonium acetate (NH_4 OAc) – Fisons anal. grade, 99%

Acetic acid (CH_3COOH) – Fisons HPLC grade

Acetonitrile (CH_3CN) – Fisons HPLC grade

Controlled pore glass (CPG) support, 15μl dC, dG – Cruachem

DNA – Grade Bio Gel hydroxylapatite – Biorad

Deuterium oxide (D_2O) – Sigma

Disodium hydrogen orthophosphate ($Na_2HPO_42H_2O$) – Fisons anal. grade, 98-100%

Sodium dihydrogen orthophosphate ($NaH_2PO_42H_2O$) – Fisons anal. Grade, 99-101%

Sodium chloride (NaCl) – BDH Chemicals AnalaR®, 99·9%

methanol (CH_3OH) – Fisons HPLC grade

Dimethylformamide-d_7 – Sigma

Adozelesin – A gift from The Upjohn Company, Kalamazoo, Michigan 49001, USA

Cyanomorpholinoadriamycin – A gift from E. M. Acton, SRI INT, Bioorg Chemlab, Menlo Pk. Ca. 94025 USA

Sibiromycin – A gift from Prof. David E. Thurston, Department of Pharmacy, University of Portsmouth, Portsmouth, Hants, U.K.

DSB-120 – A gift from Prof. David E. Thurston, Department of Pharmacy, University of Portsmouth, Portsmouth, Hants, U.K.

2.2 Oligonucleotide synthesis

An Applied Biosystems Model 381A DNA Synthesizer was used for the initial 15μl synthesis of the various oligonucleotide sequences; 5′-d(GTTCCATGGAAC)-3′, 5′-d(CICGAATTCICG)-3′, 5′d(GCAGAATTCTIC)-3′, 5′d(GTGATGAG)-3′, 5′d(GCTAGCTAIC)-3′, 5′d(GTTGCATGCAAC)-3′, 5′(GAGATCTC)-3′ and 5′d-(CGAAAAACGG)-3′.

The CPG silica support was derivatized with a nucleoside corresponding to the base at the 3′-hydroxyl end of the required DNA strand. This provided the starting material to which the dA, dT, dG, dC and dI β-cyanoethyl phosphoramidites were sequentially added *via* a 62 step synthetic cycle, as proposed by Perkin Elmer (Appendix I). The β-cyanoethyl phosphoramidites (1 g each of dA, dT, dG, dC or 0·5 g of dI) were made up to 0·1M by adding the corresponding volumes of phosphoramidite diluent *i.e.* dry acetonitrile.

β-cyanoethyl phosphoramidite	Volume of acetonitrile (ml)
A	11.2
G	11.6
C	11.8
T	13.2
I	6.8

Table 1 The volumes of dry acetonitrile required to solvate the β-cyanoethyl phosphoramidites.

An explanation of the four stages of oligonucleotide synthesis performed by the DNA synthesizer can be found in Appendix II. The addition of each deoxynucleotide was monitored by separate collection of the trityl cation which is bright orange in solution. Failure of the production of this ion signalled that the synthesis was unsuccessful.

2.3 Post-synthetic treatment of DNA

This encompasses the cleavage of the DNA from the column, deprotection of the β-cyanoethyl phosphoramidites, benzoyl and isobutyryl groups, followed by detritylation and separation of the desired strand from the failed syntheses (Applied Biosystems Manual; Thompson and Hurley (1995)).

1) Cleavage from the silica support

This was achieved by a 20-25min treatment with 5ml aqueous ammonia (NH_4OH), during which time the aqueous ammonia was repeatedly flushed through the column between two syringes. The NH_4OH cleaves the ester linkage between the CPG support and the 3′-hydroxyl of the initial nucleoside (base and sugar, no phosphate group), revealing the free 3′-OH group of the DNA strand (Fig. 38).

2) Deprotection of the phosphates

These β-cyanoethyl protecting groups are removed at the same time as the DNA is cleaved from the support, in the above reaction (Fig. 38).

3) Deprotection of the bases

This step was achieved *via* ammoniolysis and resulted in the removal of the benzoyl group (Bz) from the exocyclic amines of adenosine (A^{Bz}) and cytosine (C^{Bz}) and the isobutyryl group (Ib) from the guanosine (G^{Ib}). Thymidine, already being unreactive, did not require a protecting group. Ammoniolysis was performed by incubating a vial (which could withstand positive pressure), containing the DNA product plus failed syntheses at 40°C for approximately 15 hours (Fig. 38).

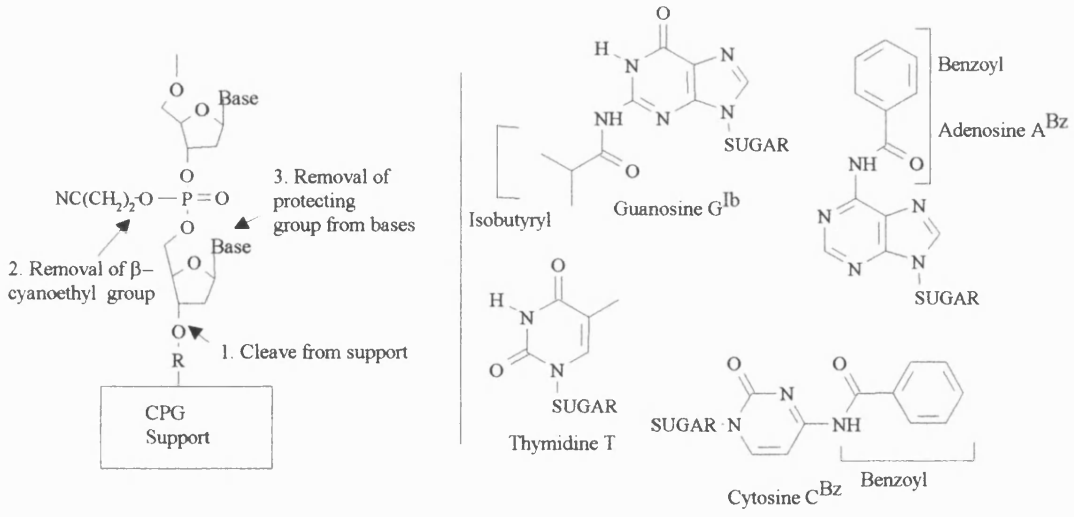


Figure 38 Cleavage from the CPG support and deprotection of phosphates and bases.

At this point, the DNA product and failed syntheses are completely deprotected except for the dimethoxytrityl group on the 5′-end of the DNA. The mixture can then be separated by HPLC.

2·4 High Pressure Liquid Chromatagraphy (HPLC)

The HPLC analysis was carrried out using a Millipore Waters 600E system controller and 490E programmable wavelength detector. Purification of the single stranded DNA was achieved using the reversed phase method with a Dynamax–300A, C_{18} (20mm x 200μm) preparative HPLC column. The DNA was injected onto the column in 100mM ammonium acetate (aq), pH 7, and then eluted (flow rate 10ml/min), using a programmed elution gradient to achieve a linear change to 30% HPLC grade acetonitrile over a period of 40 minutes (Thompson and Hurley, 1995). Each fraction correlating to the various peaks was collected independently. After this initial slow increase, a more rapid change to 100% acetonitrile was required to flush the column of any residual compounds. Polarity and molecular size form the basis of separation with small, highly polar analytes being eluted first (*e.g.* failed syntheses) and large, non-polar molecules being retained the longest.

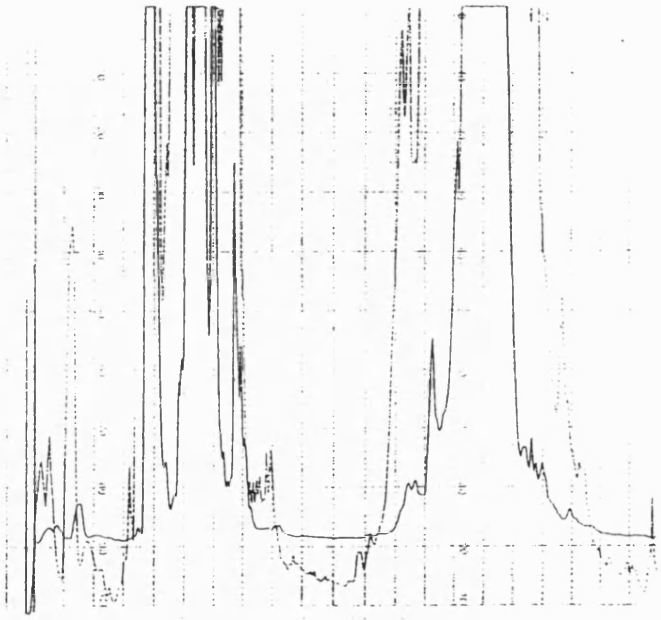


Figure 39 HPLC trace for 'trityl on' DNA

HPLC of the tritylated DNA sample produced a trace (Fig. 39), with a typical pattern, independent of the base make-up of the DNA sequence. The first set of peaks represent the 'failed syntheses' and the second large, off-scale peak corresponds to the DNA with the trityl protecting group in place. The first set of peaks corresponding to 'failed syntheses', related to the small number of DNA strands which were capped at each stage, when they did not succeed in reacting with the incoming β-cyanoethyl phosphoramidite base. The smallest molecular weight single strands were eluted first which related to just one or two bases. Longer strands of

up to n-1 nucleotides (where n = number of nucleotides in the oligomer), were consecutively eluted showing a succession of peaks. The large, separate, off-scale absorbance peak is eluted last, and corresponds to the desired synthetic product, the single stranded DNA 'n–mer'. There is a large space between the 'failed syntheses' and the product peak, which would not occur if the separation was motivated by molecular size alone. This gap is connected to the polarity of the molecules. Capping during synthesis to terminate any unreacted chains results in an acetylated 5′-OH as opposed to the dimethoxytrityl group, which occupies this position in the completed DNA strand. It is the polarity of these two groups which is responsible for their peak separation, the smaller acetyl group being more polar than the dimethoxytrityl group. Thus acetylated strands are partitioned into the acetonitrile mobile phase at a lower percentage concentration than those with the dimethoxytrityl group. Therefore, this polarity difference combined with molecular size accounts for the early elution of the 'failed syntheses' from the column, relative to the larger, less polar dimethoxytritylated DNA n-mer.

HPLC yielded a clear liquid containing the tritylated product. This sample was evaporated to approximately 5ml volume, followed by dialysis at 5°C using Spectra/PorCE membrane (MWCO:1000) against 3 x 1000ml of MilliQ water, to remove the ammonium acetate. The contents of the tubing were then concentrated by rotary evaporation, before deprotection of the 5′-OH by treating with 60ml of 50% (v/v) glacial acetic acid for 20mins at room temperature. The acid was then removed by evaporation.

The DNA sample was then ready for further purification by HPLC which was performed in the same manner as before, except an acetonitrile gradient of 0-15% over 40mins was used. This 'trityl off' fraction was collected at *ca.* 12% acetonitrile before concentration and dialysis as previously described.

For each oligonucleotide strand, NMR buffer (0·6ml) was required to add to the sample. This was prepared by mixing 0·01M NaH_2PO_4 and 0·01M Na_2HPO_4 (each made up using 0·1M NaCl) till they reached pH 6·85 (Thompson *et al.*, 1995). The DNA sample was evaporated to dryness, before being dissolved in 1ml D_2O, freeze-dried to dryness once again, and then taken up in 0·6ml D_2O. This DNA sample was placed in an NMR tube and examined by one-dimensional ^1H NMR.

Salt concentration has a significant effect on spectral resolution and so its addition prior to NMR is essential in order to obtain maximal information about the sample. Phosphate buffer maintains the pH but its own protons are likely to give rise to ^1H NMR signals larger than those of the macromolecule of interest, possibly obscuring a region of the spectrum. However, they can be exchanged by redissolution in D_2O. The phosphate buffer is important for

sustaining the three-dimensional duplex structure of the DNA, to avoid the formation of hairpins at very low ionic strengths (Primrose, 1993).

2.5 One-dimensional ^{1}H NMR spectrum

One-dimensional ^{1}H NMR data were collected on each purified oligomer, in deuterated water, using a Jeol EX400 (400 MHz) spectrometer. There were four basic procedures to be observed (Frenkiel, 1993) in setting up the instrument, before the 1D ^{1}H NMR spectrum could be acquired. These are listed as follows;

1) To tune the probe.
2) To establish the field and frequency lock.
3) Optimize the magnetic field homogeneity.
4) Calibrate the radiofrequency pulse lengths.

An explanation of these processes and how they were achieved is given below.

2.5.1 Establishing the field and frequency lock

Since electromagnets do not produce inherently stable magnetic fields, there needs to be a control mechanism to produce high resolution data. This is accomplished by regulating the magnetic field so that the deuterium resonance of the solvent is resonant with the radio frequency of the spectrometer. After this lock has been achieved, its phase is altered to give the highest signal, which is important for shimming and ensuring the signal is in pure dispersion. The lock transmitter is not run at very high power, because this could lead to detrimental effects on field stabilization and hence shimming.

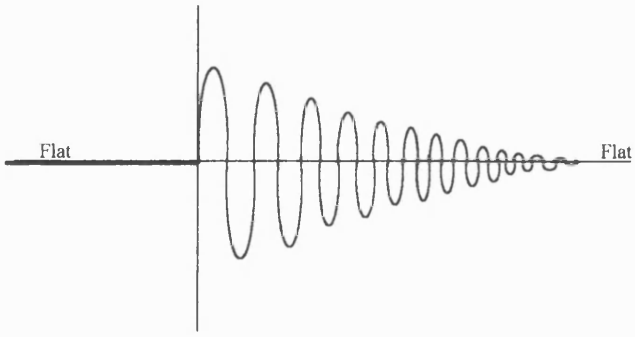


Figure 40 The Free Induction Decay (FID)

The experimental procedure to attain this field and frequency lock on the Jeol EX400 spectrometer began by reading in the ^{1}H menu (data collection between -1·5 and 15ppm) and

loading the solvent (D_2O). The lock signal was located and the FID sine wave observed to see if it was phased correctly (Fig. 40).

2.5.2 Tuning the probe

Factors such as volume, position, solvent and ionic strength of the sample can strongly affect tuning of the observation coils within the probe. Whereas the lock channel tuning is sample-independent, the probe is not and so requires tuning to increase sensitivity and produce the strongest possible radio frequency, by utilizing the transmitter to produce a low level signal at the observation frequency to which it tunes. At this point, all of the radio frequency power will be absorbed by the probe; the meter gives a measure of the reflected power which should be minimal after tuning. The probe was tuned whilst the sample was not spinning, because tuning the probe is difficult if the signal is oscillating.

2.5.3 Optimizing the magnetic field homogeneity

This is also known as 'shimming'. By varying the currents in a set of correction coils which are each designed to counteract a particular type of field gradient, the homogeneity of the magnetic field to which the sample is exposed can be improved. The lineshape is important during shimming, because if the solvent peak base ends up particularly wide, solute peaks in the vicinity may be eclipsed by it. The height of the lock signal was used as an indicator of magnetic field homogeneity during shimming. Where just a straightforward one-dimensional ^1H NMR spectrum was required, the experiment was performed with the sample spinning, so shimming also took place whilst the sample was spinning. The sample was spun to increase the resolution because the sample tube is not homologous throughout, so spinning averaged the magnetic field experienced by the sample.

2.5.4 Calibrating the radiofrequency pulse lengths

This procedure is most significant for multiple pulse experiments where the sample would be subjected to 90° or 180° pulses. For the simple 1D accumulation this was not performed.

Following the set-up process of the DNA sample in the spectrometer, and regulation of all the instrument parameters required to produce data of good resolution, is the adjustment of the acquistion parameters. All NMR experiments were run at room temperature (21°C) because

at low temperatures (<10°C), a reduction in molecular tumbling resulting from an increase in viscosity of solution or aggregation of duplexes, may cause line broadening (Feigon *et al.*, 1983a).

2.5.5 Adjustment of acquistion parameters

a) The filter was calculated make it easier to phase the spectrum.
b) The spectral width was chosen (FR 8000 although 3500 would incorporate all of the spectrum) A frequency range more than twice the width of the spectrum was used, as smaller ranges can have adverse effects on the baseline. The acquisition time and signal-to-noise ratio are independent of spectral width, so remain constant at larger frequencies.
c) Reciever gain. This was usually set to the highest possible value that can be achieved without causing the receiver front end or the analogue to digital converter to overload, which could result in distortion of the lineshape.
d) Accumulation. The acquisition time is the period over which a signal is accumulated. The spectrum is produced by subjecting the sample to a range of frequencies (ΔHz), and analysing the output before storing in a digital form. The signal is sampled twice every cycle or 2Δ times every second, where Δ = spectral width, which is also equivalent to the highest frequency component.

   Data was accumulated generally over 64 scans or 128 scans if the sample was slightly less concentrated.
e) Free induction decay (FID). This build up of the FID can be observed after each scan on screen. Because the FID is a measurement of intensity as a function of time, it is known as the time domain signal, which corresponds to an exponential decay curve of a sine wave. This signal is generated from the application of a pulse, at the Larmor frequency. For extremely complex macromolecules with many spectral lines, the FID is also complicated.

2.5.6 Transformation into a real spectrum

a) The interconversion of data between the time and frequency domain is a mathematical process, known as Fourier Transformation. The FID is Fourier transformed to the 'real' spectrum. Fourier transformation contains real and imaginery components, but only the 'real' constituent is used to generate the spectrum (Fig. 41).

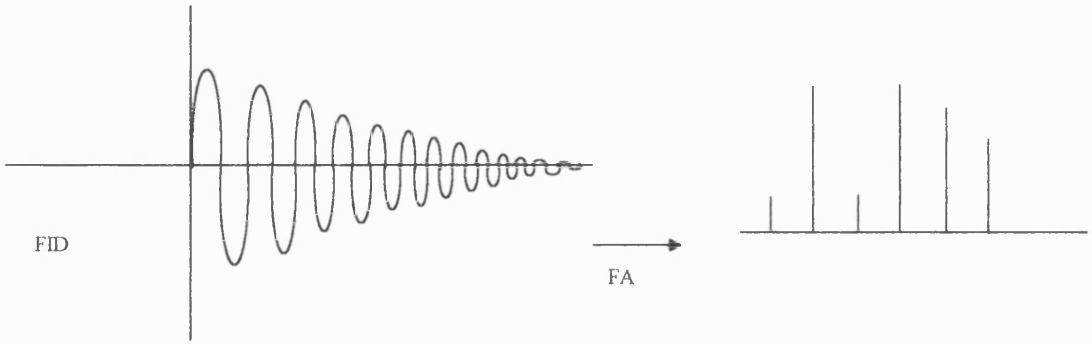


Figure 41 Fourier Transformation of the FID to a real spectrum

b) Phasing. Phasing is required to produce symmetrical peak lineshapes in the spectrum, with phasing first taking place around the largest peak *i.e.* D_2O and then at a distance from it, maximizing the area under a peak to a flat baseline. The gain along the Y axis (YG) may also be turned up at this point to increase the signal (and the noise) size.
c) Referencing. The spectrum is referenced to its largest peak, which is the solvent D_2O at 4.71ppm.
d) Data correction. Lastly, it is possible to flatten the base line to eliminate spurious noise signals and baseline roll.

This protocol produces 1D ^{1}H NMR spectra for the various DNA samples and also provides the initial set up procedure for multipulse experiments, such as NOESY and DQF-COSY, which will be discussed later.

2·6 Analysis of the 1D ^{1}H NMR spectrum.

One-dimensional ^{1}H NMR spectra were used as a tool to probe the concentration, purity and the ratio of oligomers in duplex formation from titration of non-complementary sequences, as well as the progress of reactions between the drugs and DNA. The information acquired from the spectra determined the next course of action. Line broadening (resulting from paramagnetics impurities), and the presence of organic solvent peaks indicated further purification was necessary *via* hydroxylapatite chromatography, solid phase extraction or forced dialysis to improve the appearance of the spectra. The protocols of these techniques will now be described in detail.

2.7 Annealing

For non-palindromic sequences such as 5′d(GTGATGAG)3′ and 5′d(CGAAAAACGG)3′, and their complementary oligomer strands 5′d(CTCATCAC)3′ and 5′d(CCGTTTTTCG)3′, respectively (which do not form a duplex as readily as the self-complementary sequences), annealing was necessary. This procedure also ensured removal of any intrastrand hairpin loops which may have formed, due to the end of one strand being complementary to another region and binding in the incorrect postion. This latter problem occurs in both palindromic and non-palindromic sequences. Difficulties in duplex formation lead to line broadening and poorly resolved NMR spectra.

Reannealing is not only a simple solution to this problem but is also a very simple technique. An NMR tube containing the DNA sample in question, is placed vertically in a vacuum flask with enough water to submerge the sample at a temperature of 75°C. The DNA melts (midpoint melting temperature is *ca.*69°C (Stryer, 1988)) and then begins to cool slowly back to room temperature. Annealing between two strands occurs at a higher temperature than within a strand, so all the DNA should have been incorporated into duplexes before it is cool enough to form hairpins.

2.8 Hydroxylapatite Chromatography

Hydroxylapatite chromatography was the purification technique of choice, when the DNA sample container impurities observable in the NMR spectrum. It is also an effective way of separating excess single stranded DNA from duplex DNA after the titration of two oligomers, to produce a non-complementary duplex. Hydroxylapatite has a positively charged Ca^{2+} centre, and a negatively charged phosphate centres. The principles of ion-exchange chromatography are therefore applicable; however, only some ions can elute adsorbed macromolecules. What would normally be seen as a single peak in ion-exchange chromatography or HPLC, may be separated by hydroxylapatite chromatography. Charge density, molecular weight or size, and isoelectric point are all significant factors in the separation. The underlying mechanism of this technique is not well understood. However, for a given pH, the adsorption-desorption process does seem to depend on ionic strength, with retention of DNA on the column occurring at low eluent molarity (Gorbunoff, 1984). It is important that pH 6·8 is sustained because deviation results in decreased elution of single and double stranded DNA and, therefore, decreased total recovery of DNA.

This technique was performed as follows: 100ml of dry DNA grade HTP hydroxylapatite powder (Biorad) (HA) = $[Ca_5(PO_4)3OH]_2$ was added to six parts of starting

buffer prepared from 0·01M Na_2HPO_4 and 0·01M NaH_2PO_4 mixed to pH 6·8, and then swirled gently. The slurry was allowed to settle for 15 minutes before decanting the fine particles from above the settled bed (Bio-Rad hydroxylapatite instruction manual). This process was repeated twice. After resuspension, the hydroxylapatite was poured into a Pharmacia XK26 glass chromatography column and was allowed to settle under gravity. This preceded washing with two column volumes of the starting buffer. The column was connected to a Gilson Minipulse 2 peristaltic pump flow rate 10ml/min, and a detector; Pharmacia Biotech monitor UV-1, set at wavelength 254nm, with a chart recorder and sample collection utility.

1ml of DNA solution was loaded onto the column in 0·01M phosphate buffer. Care was taken not to disturb the crystalline surface of the hydroxylapatite column which could have led to uneven adsorption of DNA onto the column. Elution took place over a gradient of 0·01M→0·15M phosphate buffer (pH 6·8), formed by the gradual mixing of 300ml portions of each buffer solution. The sample was eluted from the column in two consecutive broad bands, if separation of single and double stranded DNA was being performed. The first peak related to the small amount of single stranded DNA, whilst the second much larger peak correlated to the duplex DNA. Purification of duplex DNA only, resulted in the observation of only the second peak. Collection of this fraction gave the pure DNA duplex. Other impurities were also removed, some of which flowed straight through the column, showing a series of small peaks near the beginning of the trace being probably small and non-polar in character. The collected eluent was then concentrated by rotary evaporation and dialysed to remove the salt, before being prepared for NMR as previously described.

2.9 Solid phase extraction

Solid phase extraction is a particularly effective method of removing relatively non-polar compounds from the sample, and was used to successfully separate unreacted drug from drug-DNA adducts. This technique requires a 6ml Supelclean LC-18 (phase is octadecyl bonded silica) Sep-Pak tube, which is connected to both a reservoir and a Supelco vacuum manifold. Prior to commencing the extraction procedure, the column was conditioned with 2ml HPLC grade methanol (MeOH) to activate the packing. 10ml of 30% HPLC grade acetonitrile (MeCN) was then passed through the column in order to remove any particles originally present on the C_{18} Sep-Pak, which would be soluble at this concentration. 10ml of milliQ water was used to wash the activated column, so that organic solvents would not contaminate the DNA sample, which was then loaded. 50ml of 30% HPLC grade acetonitrile was then placed in the reservoir and allowed to drip extremely slowly through the column (vacuum did not exceed 5psi) over a period of six hours. This ensured all of the DNA was eluted from the C_{18} Sep-Pak

within the acetonitrile mobile phase. Collecting the eluent in small sample volumes and analysing the UV signature at 254nm, referenced to 30% MeCN, made it possible to discern when all the DNA had been removed from the column

The eluent containing the DNA sample was then reduced to dryness and prepared for NMR, using NMR buffer and D_2O as previously described.

2.10 Other purification techniques

HPLC, using an acetonitrile gradient of up to 15%, was also used as a mode of further purification. However hydroxylapatite chromatography and solid phase extraction produced results of far superior quality on examination of the NMR spectrum.

Freeze-drying was also a method of removing volatile materials such as methanol. Contaminants such as triethylamine were removed by forced dialysis. This involved saturating the sample solution in the dialysis tubing with NaCl to discharge the DNA, effectively allowing the charged triethylamine to be dialysed away. Contaminant paramagnetic ions were sequestered by the chelating agent EDTA, which was added (0·01M) to the 0·1M NaCl(aq) solution, used to make up the phosphate NMR buffer.

2.11 DNA adduct formation

On production of a pure DNA duplex (verified by NMR), the DNA was ready to be treated with a drug to produce a covalently bound drug-DNA adduct, or a non-covalently bound complex in the case of sibiromycin. Each DNA duplex was freeze-dried in a small vial, followed by resuspension in 500µl D_2O. The drugs were added to their corresponding binding sequences as follows:

0·5mg cyanomorpholinoadriamycin CMA(s) (red) was allowed to react with

5′d(GTTCCATGGAAC)$_2$-3′,

5′d(GCTAGCTAIC)$_2$-3′

5′d(GTTGCATGCAAC)$_2$-3′

2mg sibiromycin(s) (dull yellow) was allowed to react with:

5′d(CICGAATTCIGG)$_2$-3′

5′d(GCAGAATTCTIC)$_2$-3′

5′d(GAGATCTC)$_2$-3′

Into a vial containing 50mg of adozelesin, 250µl DMF-d_7 was added to dissolve the drug. 5µl (5mg) adozelesin (yellow) was allowed to react with:

5′d(CGTAAGCGCTTACG)$_2$-3′

5′d(CGAAAAACGG)-3′·5′d(CCGTTTTTCG)-3′

Into a vial containing 12mg of DSB120, 60µl of DMF-d_7 was added to take up the drug into solution. 20µl DSB-120 (brownish) was allowed to react with:

5′d(GTGATGAG)-3′·5′d(CTCATCAC)-3′

All solutions were then stirred in the dark at room temperature for 48 hours. The excess drug was removed from solution by vortexing and the drug-DNA adduct supernatent was decanted, taking care not to disturb the free drug pellet. The supernatent was freeze-dried to remove the solvent before further rounds of freeze-drying. After addition of NMR buffer and D_2O. The drug-DNA complex was analysed by 1D ^{1}H NMR as previously described. Low resolution spectra required further sample purification by either C_{18} Sep-Pak solid phase extraction or hydroxylapatite chromatography. The production of a clean, pure sample in high yield allowed the commencement of 2D ^{1}H NMR experimental analysis.

2.12 2D ^{1}H NMR NOESY protocol

The set-up of a NOESY experiment proceeds primarily in the same manner as the 1D experiment *i.e.* tuning the probe (which in this case is a HX probe), and obtaining a 1D reference spectrum. However, water suppression techniques employed to eradicate the residual water proton signals in the solvent D_2O peak. Solvent suppression increases the signal-to-noise ratio in the low sensitivity NOESY experiment. Presaturation is the most straightforward suppression technique and can be used on a solvent which resonates within a 20–30Hz band width. A value greater than this could result in suppression of solute peaks. Irradiation proceeds by the application of a continuous unmodulated period of low level irradiation which is applied directly at the solvent frequency (Frenkiel, 1993), prior to the NOESY pulse sequence. This results in the water z component of magnetization being reduced to zero.

The water suppression procedure was executed as follows:

1) On the processed one-dimensional ^{1}H spectrum, the water peak was identified and the spectrum zoomed to 250Hz width.
2) The water peak height was set to 80% of the screen and the apex was pin-pointed.
3) IRR, this changes the irradiation frequency (IRF). The observation frequency (OBF) value needs to be equivalent to IRF (OBF/IRF). This brings the water peak into the centre of the observation window.

4) The experiment is changed to homeogated (EX HMG SE).
5) The irratin (IRA) is set. The lower the value to which the IRA is set, the greater is the suppression of the water peak. There are 512 settings (IRA=200).
6) Once the identity of the peak to suppress and strength of suppression has been chosen, the length of time (pulse delay = PD) for which to suppress it (PD = 4), has to be elected.
7) The receiver gain is now amplified to a value of 21.
8) A spectrum is then accumulated over 4 scans and Fourier-transformed. Scrutinisation of the spectrum shows whether; the suppression has been too great (next try IRA 220) and the peaks have begun to go out of phase, the suppression was not great enough leaving some residual water (next try IRA 190), or if the water peak is now fully suppressed.
9) Zooming around the spectrum (XE 3600) exhibits the frequency around the sample (with a little space on either side). Spectral width is equivalent to 10·002ppm. The frequency is made equivalent to this value (FR 3600).
10) The filter was calculated (@FLT), to improve the phasing of the spectrum.
11) RESOL = 0·22, BF has to be half of this value *i.e.* 0·11.
12) A reference spectrum, over 16 scans is then accumulated, with phasing this time achieved using the methyl peak (next largest peak). After water supression, the FID signal appeared shorter.
13) After completion of solvent suppression, the 2D NOESY experiment was set up, using the 2D experiment menus (2DEXP @SETP).
14) The current processed 1D data were held in the memory for use in the 2D phase-sensitive NOESY experiment (0 2 VPHNOEH).
15) The mixing time was set, the smaller the molecule the larger the mixing time (*e.g.* 200, 250 or 300ms.
16) The spinner was switched to OFF, so that the sample was not spinning during the experiment.
17) The conditions were set for higher resolution (CLP 512 and TOD 512).
18) The lock signal was turned down to just below 500 (LGAIN = 16, LKLEV = 160 *i.e.* multiple of 10 between them).
19) The data were accumulated over 64 scans at each of 1024 T1 values.

The Varian 600MHz NOESY data were acquired at a mixing time 250ms with 32 scans at each of 1024 T1 values, also with a spectral width of 10·002ppm and a relaxation delay of 2s between scans, NOE spectra in 90% H_2O at 150ms mixing time were also recorded using a 1-1 echo read pulse sequence (Sklenar and Bax, 1987; Blake and Summers, 1990) with a 2s pulse repetitition delay and a sweep width of 25·025ppm.

2.13 DQF–COSY

The phase-sensitive DQF–COSY experiment is performed in a similar manner. However, the reference spectrum is not Fourier transformed, so that the FID is used. There is no mixing time and the DQF-COSY experiment VPHDQFH was selected instead of VPHNOEH.

2.14 Processing of a 2D NOESY spectrum using TRIAD

The data collected from the Jeol EX400 2D NOESY experiment was processed using TRIAD, a component of the SYBYL software suite(SYBYL 6·2 manual), on SGI R4000 and SGI R3000 Indy workstations. The first step converted the time domain raw data (.gxd and .gxp) to the FID format (.fid), *via* 'ex2nmri'. Because all the commands are initially applied to the displayed first row of the data, a method of collecting these orders is required so they can be applied simultaneously to all rows within the data set, or to another data set. The mode of data collection is a 'recipe builder'. The FID for this file was displayed half-way down the screen and on scrutinization, it was noted, that while the most decayed end of the FID was central to the mid-line, the tail end was not oscillating around zero. To resolve this problem, an average value was calculated for the last 5% of the signal and consequently subtracted from every point in the FID. This procedure is known as removing the DC offset. An apodization function *i.e.* exponential multiply was then applied and a new FID displayed on screen.

This FID time domain data was then Fourier transformed into the frequency domain, revealing an unphased 1D spectrum of negative and positive peaks. Interactive phasing produced a typical 1D spectrum of DNA or drug/DNA adduct, consisting of only positive peaks. This required adjustment of the zero order (-180-+180) and first order (-360-+360) fine and coarse controls. This concluded the processing of the first row of the first dimension. These commands were applied to the remaining rows in this dimension, followed by loading (D1 = 1 to 1 and D2 = 1 to 1024 (twice the number of points *i.e.* 512)) and displaying the subset.

Processing of the second dimension commenced with a 'data treatments' step, so that the data was rearranged prior to processing. This consisted of intermixing the 'real' and 'imaginary' elements of the complex FID vector, so that they were in the correct order for Fourier transformation. This 'data treatments' step shuffled the data sequentially, before the recipe was applied to all of the experiment in the second dimension. 'Shuffle sequential' organizes data into sequential real and imaginary data blocks. A further subset was then loaded (D1 = 1→512 D2 = 1→1) and displayed regarding the first dimension.

Two column subsets of the second dimension then required loading. These were selected manually by picking two peaks at either end of the spectrum, which were preferably isolated and of high intensity. A subset was then loaded and displayed for each of the peaks picked. Each peak was labelled by a number in points *e.g.* 242, and the data subset was loaded and named as follows; D1 = 242→242, D2 = 1→512 and the subset name was column 242. The subset for the second peaks column points *e.g.* column 458 was loaded in the same manner, before displaying each subset using the same protocol. The views were simultaneously displayed on screen using the 'sash' display option, and the recipe builder was cleared. Processing then proceeds as with the first-dimension; remove dc offset, apodize, FT and phase, where phasing adjusts the appearance of both column point peaks at the same time. This processing recipe is then applied to the whole of the experiment in the second-dimension. The last subset to be loaded and displayed, incorporates all points in each dimension of the experiment *i.e.* D1 = 1→512 and D2 = 1→512. This yields the characteristic peak contours of two-dimensional spectral data.

To optimize the appearance of the spectrum, the viewing attributes could be altered such as; number of peak contours, plotting +/- levels (only positive peaks were selected), or the lowest viewing level (this was set to just above the baseline noise). Finally, the spectrum was referenced relative to the position of the irradiated solvent D_2O peak, at a value of δ 4·71 ppm.

Some Varian Innova 600 (600 MHz) NOESY data sets were collected (courtesy of Dr. George Gray, at Varian, Palo Alto, CA., U.S.A.) and processed using TRIAD in a similar manner. The fundamental differences in the first-dimension were:

1) After removal of DC offset, the NMR calculator was utilized to multiply the first point in the vector by a half, for accurate processing in the second-dimension. The NMR calculator used the first row as the data source (row 01), was set to multiply using the 'X' option, with the real value set to 0·5 and the imaginary value set to 0·0.
2) The apodization function substituted 'exponential multiply' for 'sinebell'.
   Processing then continued as described previously for the rest of the first-dimension. There were also some changes to the protocol in the second-dimension:
3) The data treatments proceeded *via* 'shuttle interleaved', to recombine the points in the vector when the real and imaginary T_1 components are interleaved, plus reducing the vector length from 512 to 256.
4) This is followed by the NMR calculator step once again, where the data source is defined by one of the column subsets loaded in the first dimension. In this case both real and imaginary values correspond to 0·5.
5) The apodization function was again, the 90° shifted sine-bell.

6) The FID was then zero-filled to 2X, prior to Fourier transformation to give a 2X x 2X spectrum (to return the vector to 512 points).
7) After Fourier transformation the data was reversed to display the data in the usual format, with the diagonal running from the bottom left to the top right.

Processing and viewing the spectrum proceeded as before.

Jeol EX400 (400 MHz) DQF-COSY data were also accumulated and then processed in TRIAD, with a few alterations to the NOESY processing protocol. Firstly, it is important to remember the zero and first order phasing values generated whilst phasing the first-dimension. After the data treatments shuffle sequential step the recipe is not applied to the experiment in the second-dimension, instead a number of steps are jumped up to; remove DC offset, apodize *etc.* Because the FID only, is visible at the next phasing step, unlike in the NOESY experiment where two peaks were present at either end of the spectrum, interactive phasing proceeded using the values recorded from processing the first dimension. Hence, DQF-COSY spectral processing is much simpler than the protocol used for NOESY.

2.15 Building a computer model of DNA and adozelesin

Both the biological macromolecule DNA and the drug molecule adozelesin were constructed within the computer software: SYBYL. Despite the large size and complexity of the β-DNA helix, this macromolecular structure was perhaps the easiest to generate, using the in-built biopolymer dictionary facility of SYBYL. This application contains all the information required to build canonical β-DNA, using the A, G, C, T monomer structures, plus their conformations and connection rules. Two DNA duplexes were computer generated: 5′d(CGTAAGCGCTTACG)$_2$-3′ and 5′d(CGAAAAACGG)·(CCGTTTTTCG)-3′.

Adozelesin was drawn using the 'sketch' utility. Benzofuran and indole ring systems served as templates in the construction of the three sub-units. 100 Steps of steepest descent gradient minimization, followed by 10ps molecular dynamics simulation (see later protocols), were performed to find a lower energy conformation of adozelesin than the sketched one, where all rings were co-planar.

A set of charges were then generated for each molecule using the Gasteiger-Hückel method. This is a combination of the mathematical approaches described by Gasteiger and Marsili (1980) and Hückel (see Purcell and Singer, 1967 for a review). The former concentrates on the calculation of the σ component, whilst the latter on the π contribution to charge. The resulting Gasteiger-Hückel charge is the sum of these components. Because the atom charges in

the π system are believed to be delocalized over the whole π system, the Hückel charge is calculated prior to the Gasteiger-Marsili charge. The electron-distribution over the topology of the molecule is an important factor influencing the physical and chemical properties of DNA and adozelesin and, hence, the interaction between them.

2.16 Docking

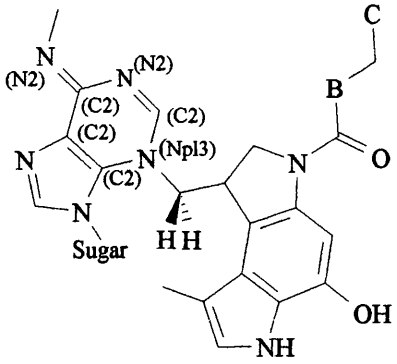
Optimal placement of adozelesin in the minor groove is of significant importance in the production of a good quality starting structure, for the initiation of a molecular modelling study. Docking was used as a tool to identify lower energy 'ligand' positions within the confined, rigid environment of the minor groove. Adozelesin was the moving 'ligand' which was manoeuvred into the stationary 'site' of the minor groove of DNA.

Before the docking procedure could commence, the centre of the DNA molecule *i.e.* the axis on which the molecule turns, had to be redefined. The adenines which were about to be covalently modified (A^{12} of the palindromic sequence and A^{7} of the non-complementary sequence), defined the new 'centre' and hence, the docking energy box encompassed this site. In the docking mode, a database was created, so as to save the docked adozelesin DNA complex as a single entity. Adozelesin was selected as the 'ligand to dock', whilst DNA was defined as the 'site to dock into' within the interactive docking mode of SYBYL. Manual ligand navigation towards the site produces a numerical total energy output as well as visual (red) lines of force. These increase in number and strength when the ligand is placed in or very close to the site, in a sterically unfavourable or electronically repulsive position. The non-bonded interactions between adozelesin and DNA are calculated on a lattice, with the co-ordinates of the ligand being converted into an index within the lattice fields. This gives a linear expression of interaction energies between the site and a particular ligand atom. When all these energies are added together for each ligand atom, the overall site-ligand interaction energy is yielded. The Tripos force field calculates steric contributions, whilst there are also electrostatic contributions from the atomic charges, both constituting the interaction energy.

The docked adozelesin-DNA complex is taken to be when the energy is lowest with few visible lines of repulsion force, and when the adozelesin appears to be in a suitable position for bonding of the CPI headunit to the adenine base.

2.17 Covalent adduct formation

Once the adozelesin had been docked into the minor groove of DNA and existed as a single entity in space, a covalent bond was added between adozelesin and N3 of adenine in DNA, using the SKETCH utility. Due to problems in locating the exact site of atomic charge on the adenine in the cyclopropapyrroloindole family of adducts, the modified adenine used in molecular modelling was as shown in Figure 42:



Modified adenine base with labelled atom types (defined by chemical elements and hybridization).

Adozelesin

Figure 42 Sketch of the modified adenine base covalently bound to the CPI headunit (Hurley *et al.*, 1984).

On covalent bonding of adozelesin to adenine in the minor groove, 200 steps of steepest descent gradient minimization was applied to a set of 'interesting' atoms; adozelesin (in both sequences), A^7, C^8, T^{14} and G^{13} of the 'A' tract sequence, or A^{12}, C^{13}, G^2 and T^3 of the self-complementary DNA strand. All other atoms were ignored, allowing the adozelesin to find a more energetically favourable position within the minor groove. This minimization was proceeded by 5ps of dynamics on adozelesin only, over a temperature gradient of 0 to 300K. The average structure from the trajectory yielded the rough starting structure for structural refinement.

2.18 The Master Assignment Spreadsheet

The master assignment spreadsheet (MSS) was generated after the adduct structure was completed, with the correct atom names and atom types as defined by SYBYL according to chemical identity and hybridization. It is used to store the chemical shift (ppm) of each atom within the covalent complex, as it is measured on the NOESY spectrum. The MSS is created from the molecular spreadsheet menu in the presence of the docked adozelesin-DNA computer model (see section 2·15). This spreadsheet contained two columns; unique atom names and shift in ppm. The number of rows correlated to the number of atoms within the complex.

2.19 The peak pick spreadsheet

Once the NOESY spectrum had been completely assigned, the NOE peak attributes such as the position in each dimension, width of peak footprint in each dimension, volume of peak integral (area under peak) and the intensity of the peak at the centre, were measured within TRIAD. The data were entered into the peak pick spreadsheet (PPSS). Each column corresponded to one of these peak attributes, whilst each row related to a peak in the NOESY spectrum. Each peak was manually picked and the peak boundaries defined. The 'add/define peak boundaries' step utilizes a rubberband tool, which is dragged out to the size of the peak in question, with its central cross corresponding to the centre of the peak. 'Selecting' the first peak automatically leads to the generation of a peak pick spreadsheet, into which all subsequent data is entered. It is extremely important to pick the exact centre of the peak, otherwise the chemical shift will be ambiguous in a spectrum where a large number of proton signals are crowded together in a small area.

The threshold level to which all the peak integrals were measured was positioned just above the baseline, to eliminate the anomalous peaks of backgound noise. This leads to a trade-off between larger volume integrals and a noise-free 'prettier' spectrum. However, the change in integral values will be proportional across the spectrum. The peak edge for defining the peak footprint size was taken to be the one closest to the peak centre (SYBYL 6·2 manual).

After all peaks corresponding to a proton-proton NOE between adozelesin-adozelesin, DNA-DNA or adozelesin-DNA had been 'picked', each proton had its chemical shift assigned (ppm). This value was entered directly into the master spreadsheet. The chemical shift of the peak in each dimension was assigned manually, this was achieved using a cross-hair tool to find the centre of the peak, which correlated with previously picked peak centres in the PPSS. At this point, it was important that both dimensions of the spectrum were referenced exactly to D_2O δ 4·71ppm. Otherwise, a proton would not have the same chemical shift value in both dimensions and problems would be encountered by TRIAD in correlating the right proton to the correct peak. A correction factor was added to the methyl group, since their single signals are equivalent to three hydrogens in rapid rotation. TRIAD was unable to generate constraints for an NOE in which three protons gave rise to the same NOE peak, leading to ambiguous assignment of each methyl proton. Therefore, the chemical shift of the methyl protons was awarded to the methyl carbon instead, and an extra 1Å added to the interatomic distance to which the NOE corresponded.

The PPSS was then linked to the MSS. All peaks (rows) in the PPSS were assigned the two protons which gave rise to the NOE cross-peak by TRIAD. The position of the cross-

peak (columns for D1 (POS 1) and D2 (POS 2)) value in ppm, was matched to the proton defined chemical shift in the MSS with a tolerance of 0·01ppm. This yielded two peak assignment columns (Atom_POS 1 and Atom_POS 2)

2.20 Generation of constraints

Distance constraints were generated in the PPSS in the presence of the adduct molecule. A strong(s)/medium(m)/weak(w) methodology was utilized (Hansen *et al.*, 1995), in which the NOE volume integrals were translated into a series of distance-range constraints. The peaks and their corresponding intensities were inspected by eye; any peaks which were completely engulfed by noise, due to the unfortunate presence of spectral baseline roll were disgarded, whilst peaks on a T_1 ridge were multiplied by 0·5, methyls divided by three, and overlaid peaks had their intensities divided by the number of protons under the peak. It was thought that total deletion of these NOEs would lead to the loss of important structural data, so intensities were weakened such that their contribution to the resulting structure was lower (weak contraints) but still present. This was thought to lead to a more accurate emplyment of the NMR data.

Generation of the distance range constraints began by selecting the assignment column (ATOM _POS 1 and ATOM _POS 2) in the PPSS. This was followed by setting the distance-ranges for the strong, medium and weak constraints, plus the peak intensity threshold;

'A' Tract;	strong	= 1·49→3·0Å	$2{\cdot}706 \times 10^8$
	medium	= 1·49→5·0Å	$1{\cdot}413 \times 10^8$
	weak	= 1·49→7·0Å	
Self-complementary;	strong	= 1.58→3.0Å	1×10^6
	medium	= 1.58→5.0Å	1×10^5
	weak	= 1.58→7.0Å	

The PPSS was updated with the distance range constraints in two columns 'LOWER' and 'UPPER' bounds. TRIAD seemed to use the molecular model to define interproton distances, so that if this distance exceeded the range, then no constraint was added. It was expected that distances here would exceed those normally seen from an NOE, purely because this was the docked adozelesin-DNA structure which had not taken part in energy refinement, and drug binding may have distorted the DNA somewhat. This could lead to unusal interproton interactions being present which would not be seen in a perfect β-helix or if the adozelesin had not been docked in the correct conformation. Therefore, each restraint 'bin' had its bounds widened to encompass the smallest distance (LOWER BOUND) and the largest distances (UPPER BOUNDS). After the first 50ps of molecular dymamics, the restraints were tightened to their correct values, where the medium UPPER BOUND was 4·5Å and the weak UPPER

BOUND was 6·0Å. The cytosine H5-H6 distance of 2·45Å (0·25nm), was used as a yardstick by which to measure all other proton-proton distances. The thymine H6-CH_3 distance of 2·9Å was used to calibrate all distances regarding a methyl group (Searle and Bicknell, 1992; Metzler *et al.*, 1990; Hansen and Hurley, 1995). Both of these constituted strong NOEs. The peak intensity threshold was set by dividing equally the difference between highest and lowest experimental NOE intensities.

2.21 Molecular modelling

To reduce the computational time required for the restrained molecular dynamics calculations, only the top seven base-pairs of the self-complementary adozelesin-DNA adduct were modelled. The whole of the 'A' tract adduct was modelled as it was non-palindromic, containing only one drug-binding site. In both cases, an additional pair of base-pairs (G-C), were added to the top and bottom of the DNA duplex, so that the 7-mer became an 11-mer, and the 10-mer 'A' tract became a 14-mer. The presence of these extra bases was to reduce terminal base-pair end fray within the calculations, whilst leaving the central 7-mer and 11-mer free from artificial restraints *e.g.* across base-pair hydrogen bonds. Slightly longer DNA duplexes are also perceived to behave more authentically like *in vivo*. However, a large trade-off has to be made with the increase in size, due to the computational limits.

In addition to the experimental NOE constraints applied to the molecule, some extra constraints were added to help stabilize the adduct during the equilibration protocol. These additional constraints consisted of:

1) Two angle constraints (defined by 3 atoms), from adozelesin C (of the H8A, H8B methylene bridge), to the N3 of adenine (to which adozelesin is covalently bound), to the C2 or C4 (of the modified adenine). Angle = 120° Force constant (fc) = 100kcal/mol.degree^2

   This is to pull the covalent bond so that it is planar to the adenine ring system.

2) Fixed distance restraints were applied across all hydrogen bonds between base pairs. The distances were set at their original values within the β-helix. Two constraints were placed between an AT base pair and three between a CG base pair. The fc for each base pair is as follows:

5′-	C^{21}	G^{22}	C^{1}	G^{2}	A^{3}	A^{4}	A^{5}	A^{6}	A^{7}	C^{8}	G^{9}	G^{10}	C^{23}	G^{23}	-3′
fc=	200	200	200	200	100	100	100	100	100	100	200	200	200	200	kcal/mol.Å^2
3′-	G^{28}	C^{27}	G^{20}	C^{19}	T^{18}	T^{17}	T^{16}	T^{15}	T^{14}	G^{13}	C^{12}	C^{11}	G^{26}	C^{25}	-5′

5′-	C^{15}	G^{16}	C^{1}	G^{2}	T^{3}	A^{4}	A^{5}	G^{6}	C^{7}	G^{17}	C^{18}	-3′
fc=	200	200	200	200	100	100	100	200	200	200	200	kcal/mol.Å^2
3′-	C^{22}	C^{21}	G^{14}	C^{13}	A^{12}	T^{11}	T^{10}	C^{9}	G^{8}	C^{20}	G^{19}	-5′

Each constraint was defined by acceptor and donor atoms of the hydrogen bond. This was to prevent melting of the DNA during the intial heating steps of the dynamics run.

3) Torsion angle constraints were added in the same manner as above, each with a fc = 50 kcal/mol.Å^2. Four atoms defined each constraint *i.e.* two either side of the hydrogen bond.

2.21.1 *In vacuo* structural refinement

A series of equilibration steps were performed on both adducts prior to commencement of the molecular dynamics calculations. This was to reduce any unfavourable steric clashes and geometric strains within the initial structure, as much as possible. The minimization procedure was as discussed below:

1) 100 steps of steepest descent minimization. (Compute→minimize)
   Termination was by a gradient – the minimization stops when the energy gradient is lower than the field value of 0·05 kcal/mol.Å
   Initial optimization was *via* 'simplex minimization' to reduce the stress of a high strain structure.
   Tripos force field and Gasteiger-Hückel charges
2) 10000 iterations of conjugate gradient minimization.
   Convergence criterion of the termination gradient = 0·01 kcal/mol.Å
3) Molecular dynamics – heating up procedure. (Compute→dynamics)
   Use same constraints as for minimization, except for the angle and torsional constraints, which were deleted.
   Tripos force field
   Heating cycle proceeded by 5ps at each of the listed temperatures: 0, 40, 80,120, 150, 180, 210, 240, 270 and 300K.
   Time step = 2fs
   Snapshot = every 100fs (where structure saved in the trajectory)
   Distance dependent dielectric function = 1
   Non-bond cut-off = 8Å
   NVT canonical ensemble = closed system with fixed size (N *i.e.* number of particles and V *i.e.* volume constant) but kept at constant temperature by contact to a heat bath.
   The initial velocity at 0K was 'Boltzman' but at subsequent temperatures it was 'Previous'.
4) Releasing the distance constraints across the hydrogen bonds between the bases.
   Temperature = 300K
   Replay above trajectory to acquire the starting structure for this step.

a) Time length = 5ps

fc = from 200→100kcal/mol.A^2 for first two base-pairs at either end of the duplex (excluding the extra CG base pairs)

from 100→75 kcal/mol.A^2 for the base pairs in the centre of the duplex between the terminals

Other parameters = as above Initial velocity = Boltzman

b) Reply above trajectory to acquire the structure for this step.

Time length 5ps

fc = of both 100 and 75→50 kcal/mol.A^2

Other parameters = as above

c) Replay above trajectory as before.

Time length = 5ps

fc = 50→25 kcal/mol.A^2

Other parameters = as above

d) Replay trajectory as before.

Time length = 60ps

fc = for;

C^1 -G^{14} and C^7-G^8 of the self-complementary adduct change the fc from 25→10kcal/mol.A^2

C^1-G^{20} and G^{10}-C^{11} of the 'A' tract adduct change the fc from 25→10 kcal/mol.A^2

All the remaining base pairs in between change the fc from 25→0 kcal/mol.A^2

Other parameters = as above

Also reload the distance range constraints from the PPSS, with the newly generated smaller range values as discussed previously.

5) The above trajectory was replayed and an average structure calculated from the last 50ps of the molecular dynamics run *i.e.* between 10ps and 60ps equal to 501 structures.

Steepest descent minimization – all other parameters as previously described.

100 steps

Covergence criteria = 0·01 kcal/mol.Å

6) Conjugate gradient minimization

10,000 steps

Convergence criteria = 0·001 kcal/mol.Å

Other parameters = as before.

This yielded the refined molecular structures for both adozelesin–'A' tract DNA and adozelesin–self-complementary duplex DNA adducts *in vacuo*.

2.21.2 *In aquo* structural refinement

This study began by adding a set of counterions (Na^+) to the adduct structure, which would mimic the effects of the NMR buffer in solution. These ions were added across each O-P-O bisector along the DNA phosphate backbone, at a distance of 6Å from the phophorus (Young *et al.* 1997). A distance range constraint from 5→7Å was also defined for each ion, with a fc value = 25 kcal/mol.$Å^2$.

The NMR data were collected on a solvated system. Hence, to calculate a structure which accurately reflects the adozelesin–DNA adduct in a biological system, the water component must also be included. This is achieved with 'explicit treatment of the solvent', where the whole system is solvated in a water droplet, allowing the solute to interact with the solvent by hydrogen bonding. Unforunately, this method is also very expensive in computational time. Molecular Silverware (Blanco, 1991) is the algorithim within SYBYL which performs droplet solvation, packing solvent water molecules successively around the solute, one shell at a time. Each solvent is added so that its van der Waals surface does not overlap with that of other water molecules. This method is preferable to others, even though it is significantly slower, because the resulting solvated system has a lower energy, making the molecular modelling equilibration protocol simpler (SYBYL manual).

To the minimized structure yielded from the *in vacuo* calculation step (2) with added Na^+ counterions, 6 layers of water (with Gasteiger-Hückel charges) were computed *via* Molecular Silverware, solvating the DNA adduct. The droplet consisted of approximately 4000 water molecules.

Equilibration and thermalization of the water molecules and Na^+ counterions, has to take place in the solvated system prior to equilibration of the DNA adduct (Westhof *et al.*, 1995; Young *et al.*, 1997; Auffinger *et al.*, 1995). This is to reduce strongly unfavourable energetic interactions between the solute and the solvent molecules, which could lead to collisions between the two and hence, irreversible deformation of the solute. The resulting molecular dynamics simulation would not reflect the system at equilibrium. The equilibration and thermalization procedure of water and Na^+ counterions, is discussed below:

A. 100 steps of steepest descent minimization
   Water only i.e. DNA and Na^+ are fixed in position (ignored)
   Termination is gradient, with a field value of 0·01 kcal/mol.Å
   Tripos force field
B. 100 steps steepest descent minimization

Water and Na^+ counterions only, DNA is fixed in position

$Na^+ \rightarrow$P-DNA fc = 25 kcal/mol.Å^2

Termination is gradient, with a field value of 0·01 kcal/mol.Å

Tripos force field

C. 100 steps conjugate gradient minimization

As above except reduce the $Na^+ \rightarrow$DNA fc to 20 kcal/mol.Å^2

D. 100 steps conjugate gradient minimization

Reduce fc to 15 kcal/mol.Å^2 for $Na^+ \rightarrow$DNA

E. 100 steps conjugate gradient miminization

Reduce fc to 10 kcal/mol.Å^2 for $Na^+ \rightarrow$DNA

F. 100 steps conjugate gradient minimization

$Na^+ \rightarrow$DNA fc = 5 kcal/molÅ^2

G. 100 steps conjugate gradient minimization

$Na^+ \rightarrow$DNA fc = 0 kcal/mol.Å^2

H. Molecular dynamics

Mobile waters only, DNA and Na^+ are fixed

Tripos force field

Dielectric function = constant

Initial velocity = Boltzman

5ps and temperature of 300K

2fs time step and snapshot every 100fs

NVT ensemble

I. Replay above trajectory and proceed with further molecular dynamics

Thermalize both water and counterions (DNA fixed) $Na^+ \rightarrow$DNA fc = 25 kcal/mol.Å^2

All parameters same as above except, 2ps spent at each of the following temperatures; 0, 50, 100, 150, 200, 250 and 300K

Initial velocity of first step is 'Boltzman', ensuing steps are 'Previous'

J. Replay above trajectory

Thermalize water and counterions (DNA fixed)

$Na^+ \rightarrow$DNA fc = 20 kcal/mol.Å^2

1ps at 300K Initial velocity = Boltzman

K. Replay above trajectory

Thermalize water and counterions (DNA fixed) for 1ps at 300K

$Na^+ \rightarrow$DNA fc = 10 kcal/mol.Å^2

Initial velocity = Boltzman

L. Replay above trajectory

Thermalize water and counterions (DNA fixed) for 1ps at 300K

$Na^+ \rightarrow$DNA fc = 0 kcal/mol.Å^2. Initial velocity = Boltzman

M. Replay above trajectory

Thermalization of the water molecules is now complete, the 'heating' procedure on the whole system *i.e.* water, Na^+ and DNA now begins. The protocol is generally analogous to that described in the *in vacuo* calculation, commencing at step (3). Only the changes to each step will be listed below:

3) Distance dependent dielectric function = constant (all dynamics steps in the *in aquo* calculation)

$Na^+ \rightarrow$DNA fc = 25 kcal/mol.Å^2

4a) $Na^+ \rightarrow$DNA fc = 20 kcal/mol.Å^2

4b) $Na^+ \rightarrow$DNA fc = 10 kcal/mol.Å^2

4c) $Na^+ \rightarrow$DNA fc = 5 kcal/mol.Å^2

4d) $Na^+ \rightarrow$DNA fc = 0 kcal/mol.Å^2

5) For the steepest descent minimization all waters were removed to reduce computational time.

6) The conjugate gradient minimization was also completed without the presence of water.

This yielded the refined molecular structures for both adozelesin –'A' tract DNA and adozelesin–self-complementary duplex DNA adduct *in aquo*.

CHAPTER 3: RESULTS AND DISCUSSION

THE 5′-d(CGTAAGCGCTTA*CG)$_2$- ADOZELESIN ADDUCT

Adozelesin is a synthetic analogue of the cyclopropypyrroloindole (+)CC-1065. Sequencing studies have shown that adozelesin, like all the mono-functional (CPI) based antitumour antibiotics has a sequence preference for 5′-TTA* (where * indicates the covalently modified base) Yoon and Lee, (1998). Hurley *et al.*, (1984) established that (+)CC-1065 covalently binds to the N3 of adenine *via* C8 on the cyclopropane ring opening. NMR and molecular modelling studies (Lin *et al.*, see review by Hurley, 1993) have already shown (+)CC-1065 binds within the minor groove spanning a total of 5 bp. Adozelesin is structurally similar to (+)CC-1065 minus the ethylene bridges of the non-alkylating sub-units, which maybe responsible for the delayed hepatotoxicity of adozelesin (McGovren *et al.*, 1984, Warpehoski and Bradford, 1988).

The ability of the cyclopropapyrroloindole adozelesin to modify covalently the self-complementary DNA sequence 5′-d(CGTAAGCGCTTACG)$_2$ has been investigated by high-field NMR and restrained molecular modelling. It was hypothesized that adozelesin would modify the 5′-TTA* consensus sequence to form a bisadduct. The structure was also expected not to display the characteristic winding of (+)CC-1065, if the ethylene bridges were in fact responsible for this phenomenon. This site contained an unfavourable GC within the flanking sequence to the 5′-side of 5′-TTA*, so it was hoped that if adozelesin did bind to this sequence it would be possible to assess what effect the GC base-pairs had on the conformation of the adduct and more importantly how this sequence selectivity was achieved.

The production of a refined molecular model of the adozelesin-5′-d(CGTAAGCGCTTACG)$_2$ adduct (Fig. 43) required a long sequential chain of events. Firstly, the DNA oligomer had to be synthesized, purified and analysed by 2D ^{1}H NMR. Secondly, after reaction of adozelesin with DNA followed by further purification, a second 2D NMR data set was collected and analysed. Lastly, distance-range constraints were generated and applied to a computerized model of the structure, by molecular mechanics and dynamics calculations. It was then possible to examine the structure to see what effects the binding of adozelesin had on DNA.

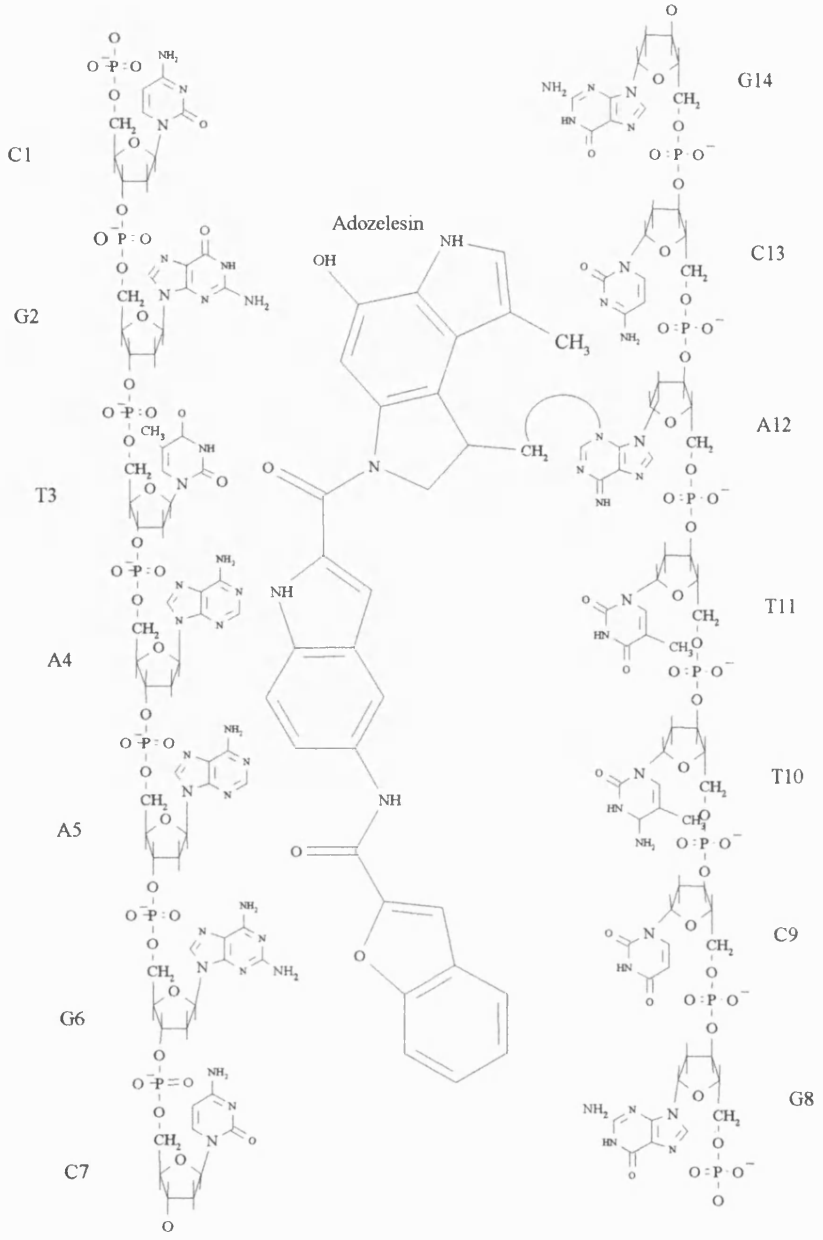


Figure 43 Schematic diagram of the adozelesin-DNA adduct.

3.1 Analysis of the 1D ^{1}H NMR spectrum

In the first instance, prior to the reaction of the oligomer with adozelesin, all non-exchangeable protons within the DNA duplex had to be assigned. This was important for later comparisons with the proton chemical shifts from the adduct structure, in order to ascertain exactly how the binding of adozelesin influences the conformation of DNA. Analysis of this spectrum also gave a strong indication on the purity of the sample. Line broadening, impurities and a low signal-to-noise ratio would constitute the main reasons for further sample purification or synthesis.

The proton numbering and distribution within the deoxyribose sugar and around each base is given in Figure 44. Although the deoxyribose protons are generally denoted as H2′ and H2′′, for the purpose of this study they will be denoted as H2′1 and H2′2, so they are equivalent to the SYBYL software naming system. The same applies to H5′ and H5′′, which will be designated as H5′1 and H5′2, respectively.

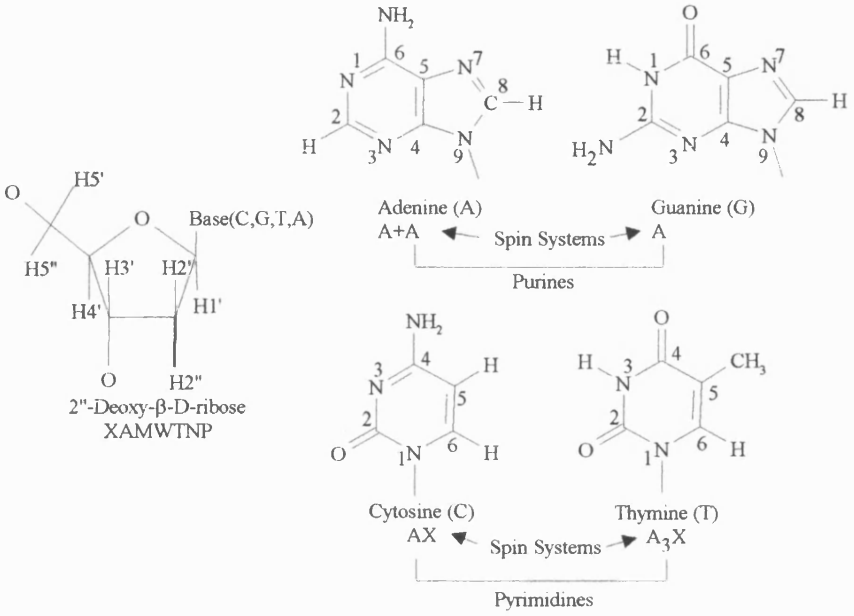


Figure 44 Molecular structures and their spin systems, of the building blocks which comprise DNA

Figure 45 shows the 1D ^{1}H NMR spectrum of 5′-d($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12}C^{13}G^{14}$)$_2$3′. Although only a limited amount of data can be extracted on the structure of the DNA duplex, it is an important tool for determining whether the correct nucleotide components are in fact present, even if their order cannot be deciphered. This is achieved by analysing the base protons of each nucleotide. The peaks due to the protons of the 2′′-deoxy-β-D-ribose sugar itself, were distributed over two-thirds of the spectrum between δ 1·8→6·0ppm. However, the sugar protons exhibited extensive overlap, so little could be deduced other than relating each spectral region to its respective deoxyribose proton: (Wüthrich, 1986) (Table 2). Due to the self-complementary nature of this duplex sequence, which means that each strand reads the same in the 5′→3′ direction, both strands are equivalent. This reduces the complexity of the spectrum significantly, because, in effect, the spectrum only carries proton resonances synonymous with one DNA strand.

Proton	Chemical shift (ppm)
CH_3	1.2 ↔ 1.6
H2′1, H2′2	1.8 ↔ 3.0
H4′, H5′1, H5′2	3.7 ↔ 4.5
H3′	3.4 ↔ 5.2
H1′	5.3 ↔ 6.3
CH5	5.3 ↔ 6.0
C/T H6	7.1 ↔ 7.6
A/G H8	7.3 ↔ 8.4
AH2	7.3 ↔ 8.4

Table 2 The chemical shift values (δ in ppm) of deoxyribonucleotide base and sugar non-exchangable protons.

From inspection of the above molecular structures, it is clear that the purines adenine and guanine give rise to two and one singlet respectively from their non-labile protons. The spin systems of cytosine and thymine are slightly more complicated. Cytosine produces doublets at both CH5 and CH6 from the ^3J scalar coupling, whilst thymine yields a methyl peak (integral of three protons) with a longer range ^4J scalar coupling to CH6, whose proton appears as a singlet in the 1D spectrum. Counting the number of methyls (unless they are overlaid) present between δ 1→2ppm gave a measure of whether or not the expected number of thymines were present in the DNA duplex. There are methyl peaks at approximately δ 1·4, δ 1·5 and δ 1·6ppm, but it is not possible to distinguish which is T^3, T^{10}, or T^{11}. In a well-resolved spectrum with a high signal-to-noise ratio, it is possible to see cytosine doublets, both in the aromatic region, just over δ 7ppm and between δ 5·3→6·0ppm for the CH5 protons. In Figure 43, only three CH5 doublets appear clearly at about δ 5·0, δ 5·2 and δ 5·25ppm, the fourth cytosine being obscured by the surrounding sugar H1′ protons. Doublets in the aromatic region, corresponding to the CH6 proton, were noted at approximately δ 7·1, δ 7·2 and δ 7·5ppm; the fourth resonance was once again overlaid with other aromatic protons within the region. Even in this one-dimensional spectrum, a sure identification of these peaks correlating to the cytosines was possible, not only because these were the peaks in the aromatic region generally with the higher upfield shifts, consistent with being a pyrimidine, but the $J_{H5\text{-}H6}$ coupling constants for all peaks were equivalent (10·8Hz), implying that the H5 and H6 were *cis* to one another. An estimation of the number of peaks in the aromatic region yielded fifteen peaks, four of which were cytosine H5 doublets, and two peaks which integrated for two and three protons. These data agreed with the expected number of aromatic resonances to be generated from the 5′-d(CGTAAGCGCTTACG)$_2$ sequence *i.e.* 4 x GH8, 3 x TH6, 3 x AH8, 3 x AH2 and 4 x CH6 (doublets) = 17 protons.

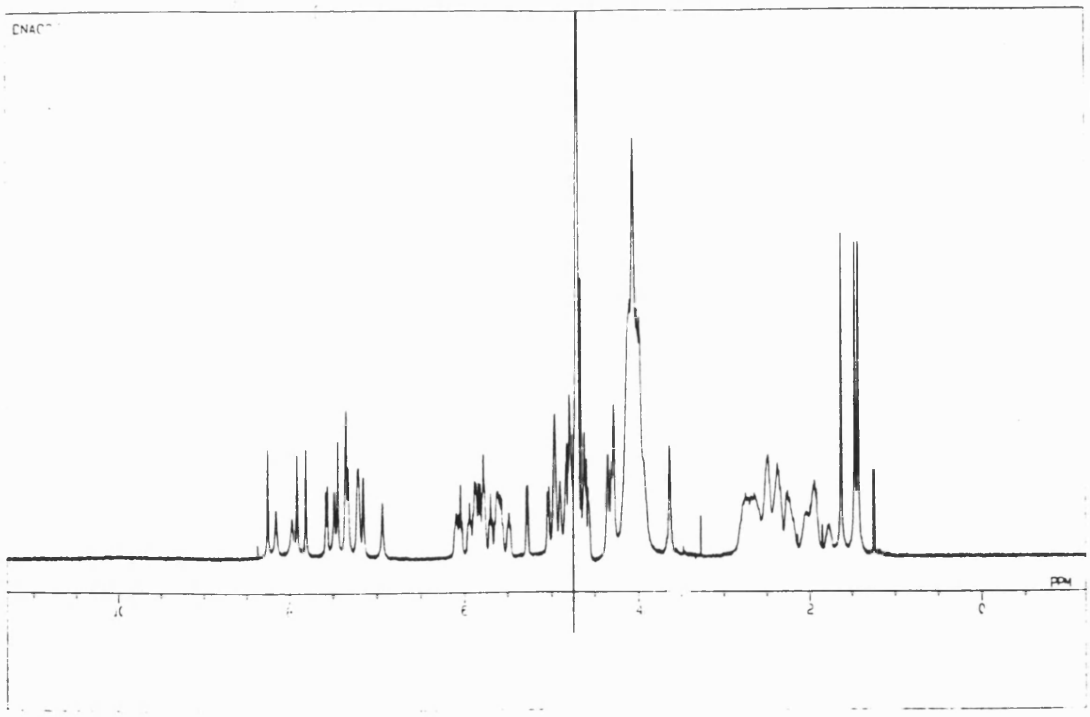


Figure 45 The 1D ^{1}H NMR spectrum of the 5′-d(CGTAAGCGCTTACG)$_2$ DNA duplex.

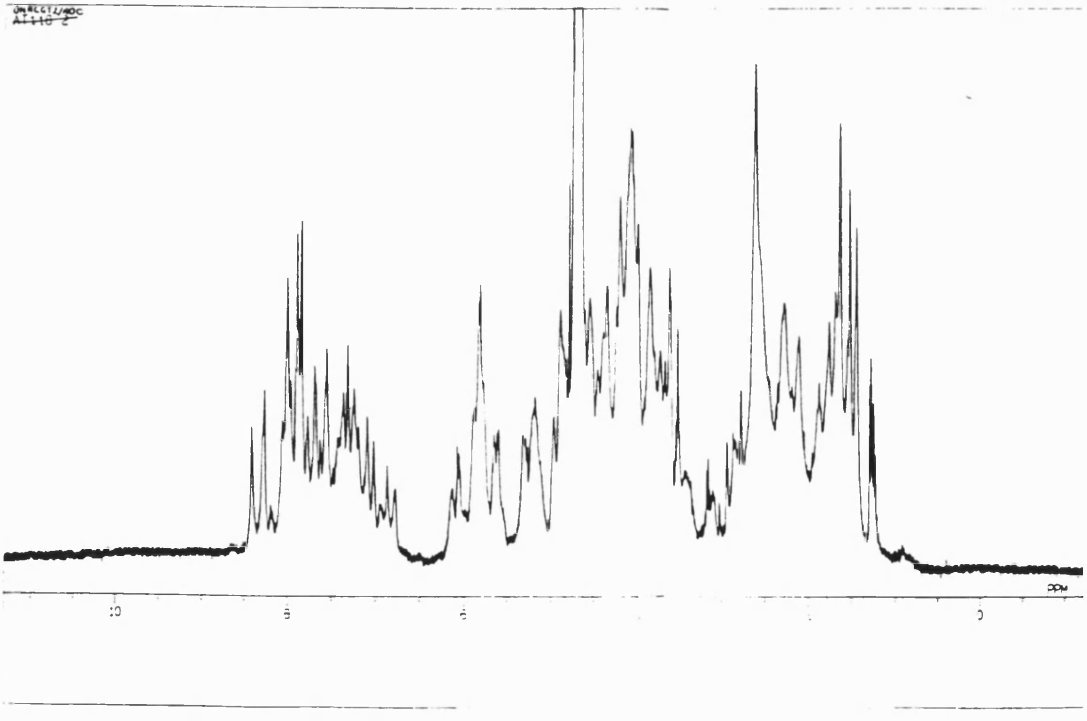


Figure 46 The 1D ^{1}H NMR spectrum of the adozelesin-5′-d (CGTAAGCGCTTA*CG)$_2$ adduct, where * denotes the site of covalent modification.

3.2 The 2D ^{1}H NOESY spectrum

After collection and processing of the Jeol 400MHz NOESY data, the spectrum was ready for assignment (Appendix III Fig. 103). The NOE peaks were well resolved, with a high signal-to-noise ratio. Each dimension was referenced to the D_2O peak at δ 4·71ppm. There were a few sharp vertical T_1 ridges of spurious noise in the spectrum at the positions of particularly strong cross-peaks on the diagonal. These were a result of systematic errors in the discrete FT; however, these did not affect spectral assignment. The NOE-observable ^{1}H-^{1}H distances were limited to a maximum distance of about 5Å. This meant that only NOEs between protons within a nucleotide (intranucleotide NOE) or between neighbouring nucleotides in the same strand (internucleotide NOE) were close enough to give rise to a through-space interaction. Each oligonucleotide strand, hence, gives rise to a cross-connectivity network between the ribose and base protons of the constituent nucleotides.

In this case, because the sequence is self-complementary, these cross-connectivity networks are equivalent and so are superimposed on one another, simplifying the spectrum. The one-dimensional spectrum previously described crosses the spectrum diagonally from the top right (δ 0ppm), to the bottom left (δ 9ppm). On each side of this diagonal is a symmetrical arrangement of cross-peaks resulting from dipole-dipole induced cross-relaxation. An NOE between two protons close in space results in a contoured peak, the x and y co-ordinates of which, when they are reverted back to the diagonal, indicate the chemical shifts of the protons. The 2D NOESY assignment procedure has been documented well (Wüthrich, 1986; Wijmenga *et al.*, 1993; Hare *et al.*, 1983; Chary *et al.*, 1987; Patel *et al.*, 1986; Weiss *et al.*, 1984; Feigon *et al.*, 1982; 1983a,b; Scheek *et al.*, 1983; 1984; Chazin *et al.*, 1986). A combination of these protocols was used in assigning the NOESY spectrum.

3.2.1 Cytosine H5 protons

To begin the sequential assignment of the protons within the DNA duplex, a suitable starting point had to be found. Inspection of the 1D spectrum had already pinpointed some doublets in the region of δ 5→6ppm, which appeared to be candidates for being a CH5. So, with this information, plus the fact that this proton is very close in space to CH6 to which it is scalar coupled in the COSY spectrum, hence probably giving rise to a very intense cross-peak in the NOESY, the spectrum was analysed. Four large intense cross-peaks (Fig. 47) were identified by their many close contours at δ 5·78, δ.5·03, δ.5·20, and δ 5·26ppm.

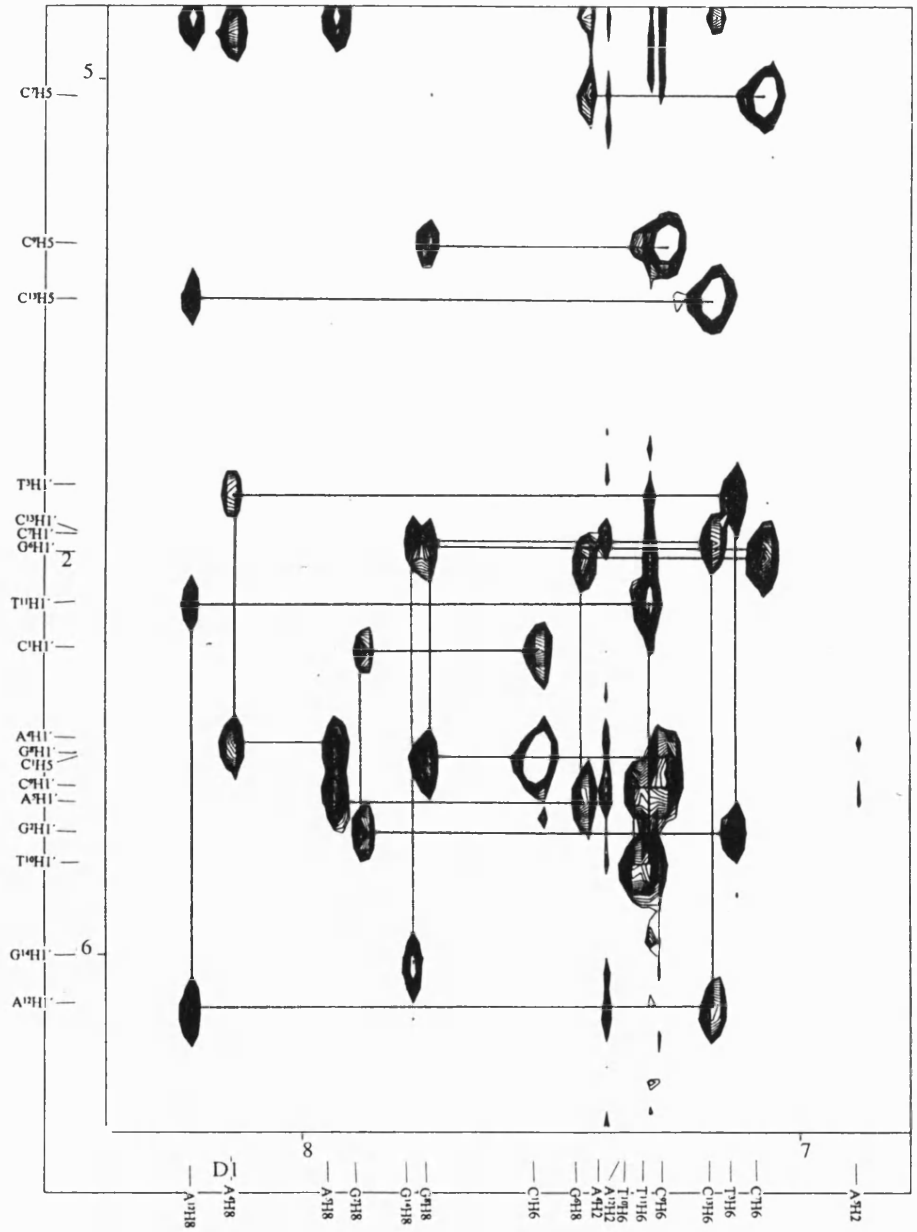


Figure 47 The 400 MHz NOESY spectrum of the H6/H8 to H1′ and H5 region.

However, these peaks could not be distinguished as being due to C^1, C^7, C^9 or C^{13} without some other identification procedure. Each CH5 peak correlated directly into its own CH1′ and CH6 protons at 4·3 and 2·4Å, respectively (Wüthrich, 1986), along the y axis (D2). However, it is also close enough to the aromatic base proton on the 5′-side (3·9Å) to produce an NOE which it connects to along the x axis (D1).

The peak at δ 5·78ppm had only one other peak it lined up with in the vertical direction, so this was assigned as the terminal because the sugar H1′ can only correlate to its own aromatic base H6 proton. At the terminal C^1, there is no 5′-base with which the C^1H1′ can interact, so this cross-peak was assigned as C^1H6→C^1H1′ and the CH5→CH6 cross-peak directly in line with it was attributed to C^1H5. The CH5 peak at δ 5·26ppm correlated with a lone peak on the D1 axis, which was the furthest downfield peak in the aromatic region and was hence most likely to be that of an adenine. Only C^{13}H5 has an adenine as the flanking 5′-nucleotide base, so this CH5 cross-peak was attributed to C^{13}H6→C^{13}H5 and the adenine H8 correlated with it was assigned as A^{12}H8. Assigning the cross-peaks at δ 5·03 and δ 5·20ppm as C^7H5 or C^9H5 proved more difficult, because each was next to a 5′-guanine. Analysis of other areas of the spectrum was therefore required to distinguish these protons.

3.2.2 Thymine CH_3 protons

The methyl protons produce a set of very intense cross-peaks in a discrete region furthest upfield between δ 1 and δ 2ppm (Fig. 48). Looking down the D2 axis, there are three rows of peaks corresponding to the T^3, T^{10} and T^{11} methyls, each row consisting of a pair of peaks. One peak is particularly intense and this is consistent with the short intranucleotide distance (2·9Å) between TCH_3→TH6 (Wüthrich, 1986). By analogy to the cytosine H5, the second peak is consistent with an NOE to the 5′-flanking, aromatic base proton. This NOE is slightly less intense, due to the greater distance between the protons (3·8Å). In each pair, the TCH_3→TH6 peak was taken to be the most intense one. Beginning with the first row, the strong upfield intranucleotide methyl cross-peak (TCH_3 δ 1·42ppm) correlates into a peak much further downfield, in the region where one would expect to find an adenine or a guanine aromatic proton. Only one thymine has a guanine to its 5′-side and this is T^3, with the neighbouring guanine being G^2, yielding an NOE from G^2H8→T^3CH_3. The remaining thymine methyls have an associated NOE with the aromatic H6/H8 proton of their neighbouring 5′-base, one of which (δ 1·63ppm) is directly in line from the D1 axis (δ 7·35ppm) with the methyl at δ 1·50ppm.

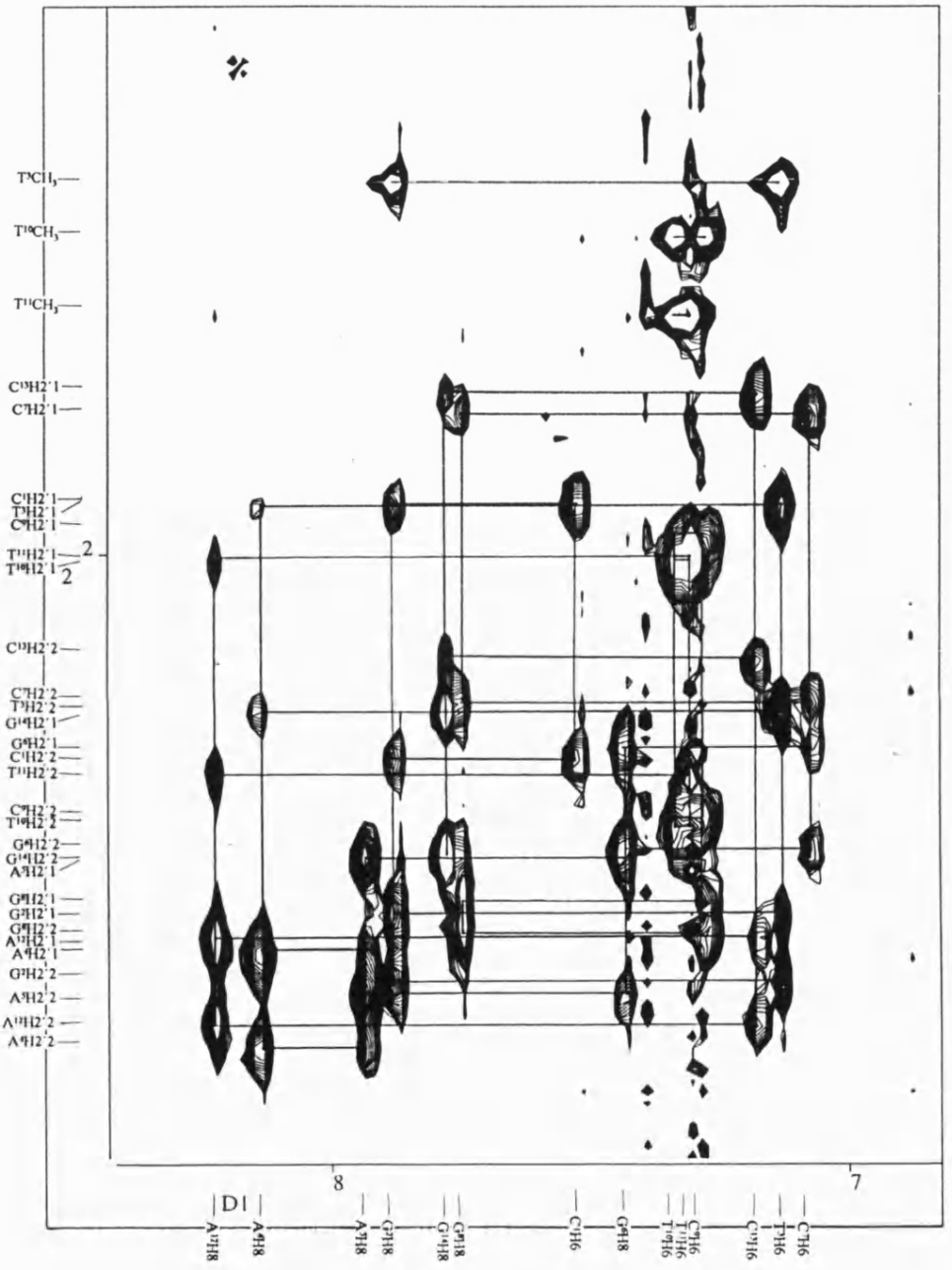


Figure 48 The 400MHz NOESY spectrum of the aromatics to sugar H2'1, H2'2 and CH_3 NOE region.

Because T^{10} is on the 5′-side of T^{11}, there will be a cross-peak $T^{10}H6 \rightarrow T^{11}CH_3$ which lines up horizontally with the $T^{11}CH_3 \rightarrow T^{11}H6$ resonance and vertically with the $T^{10}CH_3 \rightarrow T^{10}H6$ resonance, forming a small chain of connectivity. This resulted in $T^{11}CH_3$ being assigned a chemical shift of δ 1·63ppm and $T^{10}CH_3$ assigned to δ 1·50ppm. C^9 is situated on the 5′-side of T^{10} and the resulting cross-peak $C^9H6 \rightarrow T^{10}CH_3$, with which the $T^{10}CH_3 \rightarrow T^{10}H6$ peak is directly in line, also completes the assignments in the cytosine H5 region. C^9H6 has a chemical shift of δ 7·29ppm and this bisects the cytosine H5 cross-peak at δ 5·20ppm. Hence, the CH5 peak at δ 5·03ppm is assigned C^7H5 by default. Once the resonances of C^7H5 and C^9H5 had been distinguished, their 5′-flanking guanines were assigned to the connected cross-peaks *i.e.* $G^6H8 \rightarrow C^7H5$ where G^6H8 resonates at δ 7·44ppm, and $G^8H8 \rightarrow C^9H5$ where G^8H8 resonates at δ 7·754ppm.

From a complete analysis of the cytosine H5 and thymine methyl resonances in this spectrum, not only have the chemical shifts of these protons in cytosine and thymine been deduced but also those of the connected aromatic base protons. This includes intranucleotide NOEs to TH6 and CH6, as well as internucleotide NOEs to the neighbouring 5′-GH8, AH8, TH6 or CH6. Preliminary assignments of the aromatic base protons have therefore been made for:

$$5'\text{-d}(C^1G^2T^3\_\_G^6C^7G^8C^9T^{10}T^{11}A^{12}C^{13}\_)3'$$

Only three bases denoted by a blank space in the above sequence have no assignments of any proton. Analysis of these regions within the spectrum demonstrates the wealth of information that can be gained. In this case, a starting point for spectral assignment of eleven out of fourteen nucleotides in the self complementary duplex has been attained in at least one dimension.

3.2.3 The aromatic→H2′1, H2′2 ribose proton region

After the regions containing the signals from cytosine H5 and thymine methyl protons, the aromatics to H2′1, H2′2 region of the spectrum contains the next most intense set of cross-peaks. The intranucleotide distance from H8→H2′2 is 2·1Å, whilst H6→H2′2 is 2·2Å in β-DNA (Wüthrich, 1986). The H8→H2′1 intranucleotide distance is 3·8Å, whilst H6→H2′1 is 4·0Å. The very short distances to the H2′2 protons account for the high intensity of the cross-peaks. These deoxyribose protons have chemical shifts between δ 1·8 and δ 3·0ppm. Figure 48 shows that this is a crowded region of the spectrum with a high number of NOE cross-peaks between the same proton types. The terminal nucleotide was a good point to start the assignment because the base aromatic proton (C^1H6) chemical shift was already known. Moving parallel to the D2 axis, from δ 7·54ppm on D1, two cross-peaks were encountered in

this region and these were assigned as $C^1H6 \rightarrow C^1H2'1$ and $C^1H6 \rightarrow C1H2'2$. At this point, it is not possible to distinguish which cross-peak is a result of the NOE with H2′1 or H2′2. From each of these peaks, moving perpendicular to the D2 axis, the next nucleotide in the sequence was sought. The G^2H8 proton has NOE connectivities with the 5′-flanking deoxyribose H2′1 and H2′2 protons. These internucleotide distances are greater than the intranucleotide ones and thus these peaks are slightly less intense; however, they line up perfectly with $C^1H2'1$ and $C^1H2'2$. Next, in the same perpendicular direction from D2 intersecting the axis at δ 7·89ppm (the chemical shift of G^2H8), two other more intense peaks can be identified which are the intranucleotide $G^2H8 \rightarrow G^2H2'1$ and $G^2H8 \rightarrow G^2H2'2$ NOE connectivities. From aligning the base-deoxyribose internucleotide cross-peaks to find the chemical shift of G^2H8, the preliminary assignment from its connectivity with the thymine methyl has been verified, as all the cross-peaks align yielding an equivalent chemical shift value. In the same manner as the assignment procedure progressed from the terminal nucleotide to the guanine sugar protons *via* the base, sequential assignment of these protons advanced along the duplex DNA strand in the 5′→3′ direction.

Looking at the assigned H6/H8→H2′1/H2′2 region of the NOESY spectrum, it is interesting to note that the purine and pyrimidine base type is reflected in the chemical shift. Adenine and guanine ribose protons are generally shifted further downfield than those of the pyrimidines, while cytosine and thymine H2′1 and H2′2 have a greater separation in their chemical shifts than the purines. This is most likely to be a result of the stronger ring current in the purines, leading to more effective deshielding of the H2′1 and H2′2 protons and hence a downfield shift (Scheek *et al.*, 1983). As for the greater chemical shift separation of these protons in pyrimidines, this is due to the C2 carbonyl bond anisotropy, whose deshielding effect is more apparent at H2′2 than H2′1, resulting in an increased downfield shift of the H2′2 proton.

3.2.4 The aromatic→H1′ deoxyribose proton region

This region of the 2D NOESY spectrum (Fig. 47) is often assigned in conjunction with the aromatic→H2′1/H2′2 section, either to verify assignments of aromatic protons or to provide an alternative course to the connectivity pathway between nucleotides, when the cross-peaks become too overlaid in one area. The aromatic→H1′ protons are particularly well resolved in this spectrum, probably owing to the long mixing time of 300ms where spin diffusion provides a substitute cross-relaxation pathway. This allows the build-up of NOEs at increased distances, for example; H6→H1′ = 3·5Å and H8→H1′ = 3·6Å (Wüthrich, 1986),

which would normally be closer to the NOE distance limit, yielding weaker NOE cross-peaks in the spectrum which are more difficult to assign.

Examination of the cytosine H5 region which overlaps with the H1′ area has already given the chemical shift of the terminal base proton from the CH6→CH5 cross-peak. Examination of the correlations from the remaining cytosine H5 protons also yielded the H8 shift of their 5′-flanking bases. In the same manner, thymine produced base proton chemical shift information which was used in the H6/H8→H2′1/H2′2 assignments, as it will be used in the H6/H8→H1′ assignments. The more regions that are assigned, the easier it becomes to solve another, because the knowledge of chemical shifts in one dimension makes it a less difficult task to move sequentially from nucleotide to nucleotide. Beginning at the terminal C^1H6 (Fig. 45), only one cross-peak (other than CH5) is visible in the vertical plane *i.e.* to C^1H1'. G^2H8, the terminal 3′-base, is close enough to have an NOE to the 5′-flanking sugars H1′, as well as to its own H1′. Moving horizontally from $C^1H6 \rightarrow C^1H1'$, $G^2H8 \rightarrow C^1H1'$ is reached, then $G^2H8 \rightarrow G^2H1'$ is arrived at proceeding vertically. The whole procedure is then repeated sequentially until all the H1′ ribose protons have been accounted for. It is possible to start from any surely assigned proton in the strand, advancing sequentially forwards or backwards; however, the terminals in this case were a convenient starting location.

3.2.5 The H6/H8→H3′ region

H3′ protons have a chemical shift range of δ 3·2 - δ 5·2ppm. This proton 'walk' from nucleotide to nucleotide along the DNA strand is assigned using the same approach as with the H6/H8→H1′ walk. The aromatic proton chemical shifts are the same through each spectral region, so only the deoxyribose H3′ protons require identification, beginning at the terminal $C^1H6 \rightarrow C^1H3'$ at δ 4·61ppm (Appendix III Fig. 104). The longer mixing time (300ms) is extremely important in the buildup of NOEs between H6/H8→H3′, where the interproton distance is 4·6Å (Wüthrich, 1986), at the limit of NOE detection. The transfer of magnetization from, for example H6→H3′, often occurs *via* H2′2 because the H6/H8→H2′2 and H2′2→H3′ distances are considerably smaller at 2·2Å; this is, therefore, an efficient spin diffusion pathway. The NOE cross-peaks are of medium intensity which increases with the mixing time. NOESY spectra become more peak-rich and hence complex under the effects of spin diffusion; however, NOEs between all protons need to be analysed in order to provide information on the conformation of DNA. Using a longer mixing time than 250 or 300ms for a molecule of this size would produce a problem when measuring the integrals of peaks because they would become artificially high, suggesting that the interproton distance is smaller than it actually is. At this mixing time, spin diffusion is unlikely to falsify resonance assignments, as cross

relaxation is most effective around the deoxyribose within each nucleotide and less along the assignment pathways.

3.2.6 The H6/H8→H6/H8 region

This region of the spectrum (Appendix III Fig. 105) is typified by the aromatic peaks of the 1D spectrum, passing through it on the diagonal. There is one small T_2 ridge at around δ 7·3ppm, which coincides with one of the tallest peaks in the spectrum (apart from the methyl groups), where close inspection of the diagonal at this site also reveals traces of T_2 ridges. These ridges run horizontally across the spectrum, have a negative sign and cause distortions in the baseline (Wüthrich, 1986). Little can be done to rectify T_2 ridging resulting from errors in the FT and environmental conditions; fortunately, T_2 ridging is minimal in this spectrum. Cross-peaks in this area are weaker than in other regions previously described, with interproton distances of up to 5Å, spin diffusion being unlikely to progress along this route. At this point, the chemical shift of each base proton in the 5′-d(CGTAAGCGCTTACG)$_2$ DNA duplex is already known, so it is possible to jump form one base proton to another along each oligomer of DNA duplex. This results in an NOE cross-peak between $C^1H6 \rightarrow G^2H8$, $G^2H8 \rightarrow T^3H6$, $T^3H6 \rightarrow A^4H8$ *etc.* In more complex spectra, this zone is a valuable tool for finding the path from one nucleotide to another.

Using this area of the spectrum simultaneously with the aromatic to H1′ region, the adenine H2 protons can be assigned. Neighbouring adenine nucleotides *i.e.* A^4 and A^5 have an NOE connectivity between their AH2 protons, which will yield a cross-peak similar to those described above. The interproton distance of adjacent adenines is 3·6Å (Wüthrich, 1986), leading to a weak/medium intensity NOE at the coordinates A^4/A^5 δ 7·40/6·89ppm. The adenine H2 protons also tend to leave a trail down the length of the spectrum, perpendicular to D1, making them easier to spot. Using this premise the third AH2 is at δ 7·31ppm, which can also join in the $A^4 \rightarrow A^5$ base proton walk. This is because of the position of adenine H2 in the centre of the minor groove which allows it to come into close contact, not only with protons within its own strand but also to produce interstrand NOE connectivities. A^{12} is the adjacent cross-strand nucleotide to A^4 in the self-complementary duplex. An adenine H2 walk can, therefore, proceed *via* a ladder from $A^{12}H2 \rightarrow A^4H2 \rightarrow A^5H2$, cross-peak to cross-peak. Analysis of the aromatics to deoxyribose section of the spectrum was required to verify which AH2 resonance at the end of the short walk belonged to A^{12} and which to A^5. A^4H2 connected both protons (δ 7·40ppm). Each adenine H2 proton can 'see' into its own deoxyribose H1′, to the thymine H1′ to which it is Watson-Crick based paired, and to the H1′ proton of the 3′-neighbour of this thymine. In normal base-stacked β-DNA, the latter NOE connectivity is the

weakest of the three, due to the long distance. Examination of the spectrum resulted in the assignment of A^4H2 to δ 7·40ppm, A^5H2 to δ 6·89ppm and $A^{12}H2$ to δ 7·31ppm.

3.2.7 The H1′→H2′1/H2′2 region

Once the sugar H1′, H2′1 and H2′2 resonances have been assigned, it is an easy exercise to assign the NOE cross-peaks in the H1′→H2′1/H2′2 region. Beginning with the $C^1H2'1$ and $C^1H2'2$ peaks, the $C^1H2'1/H2'2$→H1′ cross peaks are sought by moving from the $C^1H2'1/H2'2$→C^1H6 peaks parallel to D1 until they both perfectly intersect a pair of cross-peaks in the H2′1/H2′2→H1′ region. Similarly, from C^1H6→H1′, by 'bouncing' off the diagonal in a perpendicular manner, this C^1H1' resonance will intersect the pair of newly found C^1H1'→$C^1H2'1/H2'2$ cross-peaks. In this way, all the other peaks in this area of the spectrum are assigned (Appendix III Fig. 106).

This is not merely an exercise in confirming that these proton chemical shifts have been correctly assigned by the presence of the relevant cross-peak. More importantly, this region holds the key to which peak of the pair corresponds to H2′1 or H2′2. The internucleotide distance between H1′ and H2′2 in deoxyribose is much shorter than that of H1′→H2′1, because they are on the same face of the sugar. This allows the assignment of the most intense cross-peak in each pair to the H2′2 proton. This DNA duplex followed the general trend that the H2′2 proton resonated further downfield than the H2′1 proton, suggesting the H2′2 proton is more deshielded by purine and pyrimidine ring currents than H2′1. Figure 107 (Appendix III) maps the H2′1 to H2′2 connectivities.

3.2.8 The H6/H8→H4′and H6/H8→H5′1/H5′2 region

This region of the 2D NOESY spectrum (Appendix III Fig. 104), which is in the range of δ 3·7 to δ 4·5ppm, is extremely crowded owing to its occupation by three sets of cross-peaks, between the deoxyribose sugar and base. Visualisation of this area of the spectrum is, once again, a result of the spin diffusion pathways around the sugar, because the H6/H8→H4′/H5′1/H5′2 interproton distances are in excess of 5Å. Shorter distances H1′/H3′→H4′/H5′1/H5′2 often produce more well defined resonances (Appendix III Fig. 108). However, peak alignment in more than one area of the spectrum is required to assign these proton chemical shifts accurately, owing to extensive peak overlay.

Considering the H6/H8→H4′ resonances first, with close correlation to the H1′→H4′ and H3′→H4′ regions whose peaks were slightly more intense, a good starting location was searched for. In an analogous manner to H6/H8→H1′ and H6/H8→H3′, a sequential walk should be able to be achieved from one end of the DNA strand to the other. In practice, due to the long internucleotide, sugar→base distances, the peaks are weaker or sometimes missing, resulting in a fragmented walk. Once again, using C^1H6, C^1H1' and C^1H3' as reference points, two peaks were noted perpendicular to D2 in these regions, one medium and one which was broader and more intense. The medium-sized peak at lower field was assigned as C^1H4' because it aligned directly with a peak at G^2H8, where this aromatic proton would be 'seeing' the sugar proton of its 5′-neighbour, as previously described. Downfield of this G^2 cross-peak is a more intense cross-peak which correlates into a weaker one at T^3H6. Assignment continued in this manner until a break in the walk emerged and, if the H1′→H4′ regions could not definitively produce an answer to the location of the H4′ for a specific nucleotide, another starting point was used. From here it was possible to advance forwards or backwards until all assignments were achieved.

The methylene H5′1 and H5′2 protons are situated between the deoxyribose ring and the phosphate group of the backbone. Looking along the O5′→C5′ bond in the direction of the mainchain, H5′1 is pro-S and H5′2 is pro-R. These geminally coupled hydrogens are not identical and therefore do not resonate at the same frequency. Their typical coupling constant is between 0 and 25Hz (Williams & Fleming, 1987). Because this region is particularly crowded with cross-peaks, although the H4′ protons have already been provisionally assigned, it is easy to confuse H4′, H5′1 and H5′2. For the nucleotides in which H5′1 and H5′2 are more separated, it is possible to observe a COSY peak situated just off the diagonal, between δ 3·7-δ 4·5ppm. These peaks are quite intense, owing to the significant amount of cross-relaxation that takes place between them and to the short inter proton distance of 1·8Å. These COSY-style peak coordinates can be aligned with the cross-peaks in the H6/H8, H1′, H3′, H4′→H5′1/H5′2 regions, to assign them to the relevant nucleotide. The H5′1 and H5′2 proton resonances were distinguished according to Remin and Shugar's rule (1972); that stated the chemical shift of H5′1 is greater than that of H5′2. The data also generally agreed with their observation that the coupling constant $J_{4'-5'1}<J_{4'-5'2}$.

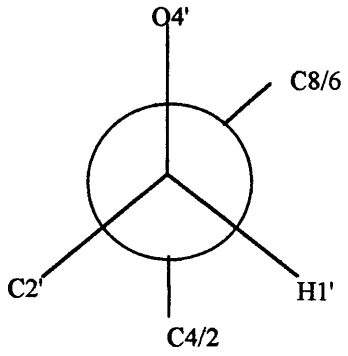
The presence of the 5′-terminal phosphate group, which can freely rotate, may explain why the $C^1H5'1$ and $C^1H5'2$ are shifted upfield in comparison to the other H5′1 and H5′2. Anisotropic effects result from the circulation of electrons around the P=O axis when it is parallel to the applied field, leading to an induced magnetic moment which affects neighbouring nuclei. Anisotropy around the P=O of the terminal phosphate, which rotates into more energetically favourable positions in solution, produces two deshielding cones extending in

either direction from the centre of this P=O double bond. The methylene protons possibly lie perpendicular to the P=O double bond axis, outside these conical areas, so they are shielded and experience a high-field shift. This effect at the terminal must override the usual deshielding effect resulting from the negative phosphate group, perhaps due to the terminals free movement in solution. The polarizing effect of the negatively charged terminal phosphate shifts the H2′2 resonance downfield and H2′1 upfield (Chazin *et al.*, 1986). However, each methylene proton of C^1 is in a similar environment, as demonstrated by their close chemical shifts.

3.3 Sequential resonance assignments summary

These 400MHz 2D ^1H NMR data clearly show that the 5′-d($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12}C^{13}G^{14}$)$_2$ duplex is β-form DNA. All sequential cross connectivity walks are present between base aromatic and deoxyribose protons, both within the nucleotide (intranucleotide) and between nucleotides (internucleotide) within the same strand, proceeding in a 5′-to-3′ direction (Fig. 50). Also contributing to the evidence for right-handed helicity is the observation that the cross-peaks relating to the methyls are all of the same relative intensity, suggesting that interproton distances are nearly equal, with the base pairs stacked evenly. The pyrimidine H5 and CH_3 protons exhibited strong NOEs to H6 or H8 of the adjacent nucleotide in the 5′-direction, as did the other deoxyribose protons in the strand, in accordance with a right handed helix. The quality of 400MHz data is not good enough to determine the exact conformation of the deoxyribose sugar geometry as endo or exo.

Feigon *et al.*, (1983b), suggested that if the H2′1→H8/H6 cross-peak was more intense than the H2′2→H8/H6 cross-peak, then the nucleotide is in anti conformation. The glycosidic torsion angle (χ) is the angle between the sugar ring and the base (Fig. 49). Where (χ) (Py) is defined by 04′-C1′-N1-N2 and χ(Pu) is defined by 04′-C1′-N9-C4 (Wijmenga *et al.*, 1993). Examination of the spectrum can produce no more than an approximation that the H2′1→H6/H8 peaks of C^1, T^3, A^4, A^5, C^7, A^{12}, C^{13} and G^{14} appear to be more intense than the H2′2→H6/H8 cross-peaks, thus suggesting the nucleotide is in the anti conformation.



χ= anti = 180° ± 90° (syn =0 ± 90°)
Typical β-DNA value= -95°
Sugar at front base at back

Figure 49 Newman projection of the glycosidic torsion angle.

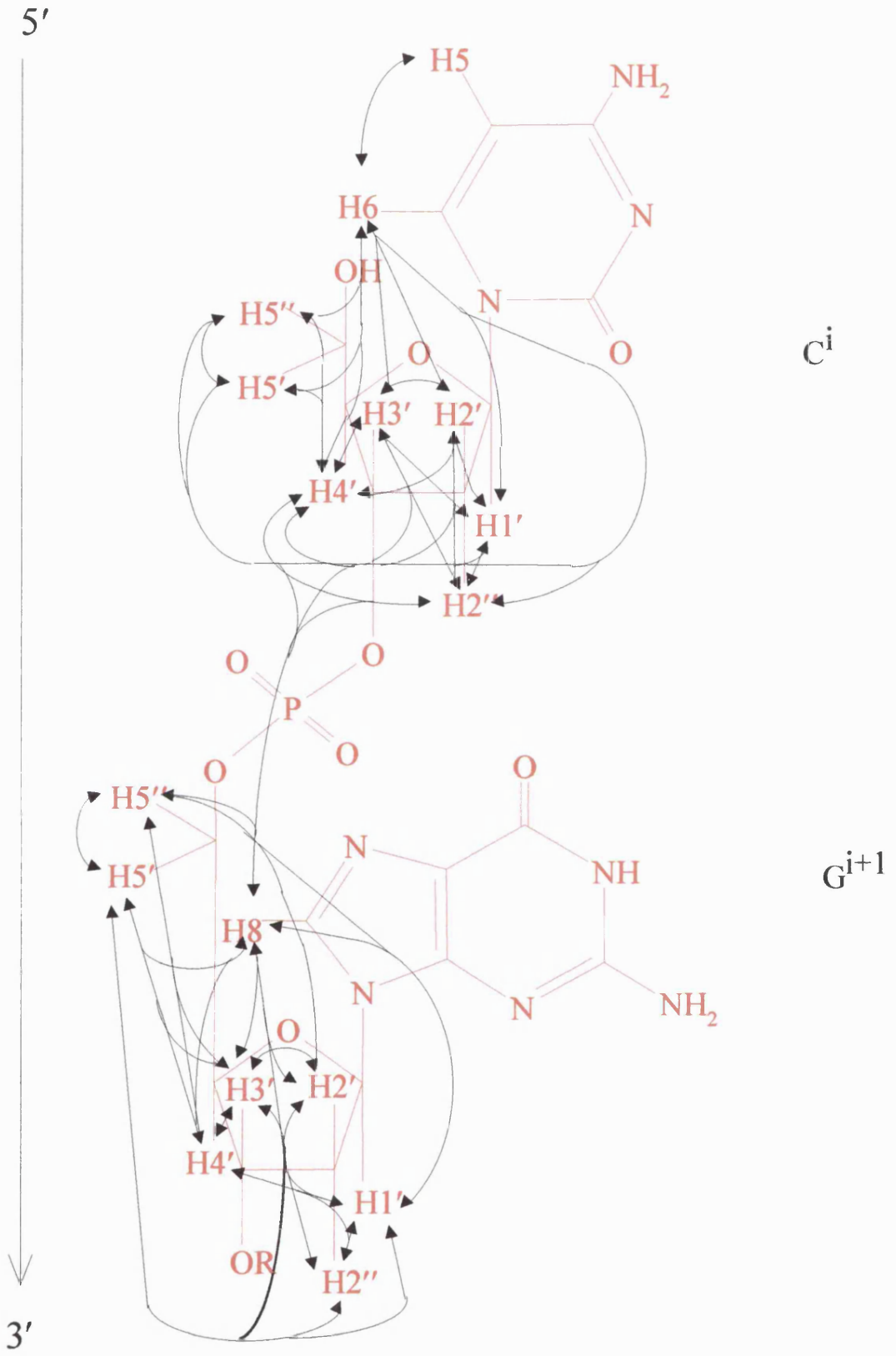


Figure 50 Schematic representation of sequential intranucleotide and internucleotide NOE connectivities (arrows), between neighbouring nucleotides in β-forn DNA as seen in the NOESY spectrum.

Table 3 Chemical shifts (ppm) for all DNA protons in the DNA duplex and adozelesin-5'-(CGTAAGCGCTTACG) DNA adduct.

	H8/H6			H2/H5/CH3			H1'			H2'			H2'2			H3'			H4'			H5'1			H5'2			Imino
C1	7,543	7,512	*-0,031*	5,784	5,744	*-0,04*	5,661	5,577	*-0,084*	1,934	1,803	*-0,131*	2,334	2,202	*-0,132*	4,608	4,567	*-0,041*	3,986	3,934	*-0,052*	3,65	3,922	*0,342*	3,623	3,547	*-0,076*	
G2	7,893	7,829	*-0,064*				5,874	5,716	*-0,158*	2,584	2,457	*-0,127*	2,692	2,508	*-0,184*	4,892	4,833	*-0,059*	4,264	3,972	*-0,292*	4,022	3,951	*-0,071*	3,937	3,791	*-0,146*	*11,45*
T3	7,137	7,33	*0,193*	1,419	1,555	*0,136*	5,485	5,135	*-0,35*	1,937	2,192	*0,255*	2,262	2,367	*0,105*	4,772	4,798	*0,026*	4,057	4,149	*0,092*	4,014	4,022	*-0,008*	3,924	3,975	*-0,051*	*13.40*
A4	8,145	8,223	*0,078*	7,401	7,809	*0,408*	5,769	6,067	*0,298*	2,642	2,574	*-0,046*	2,791	2,773	*-0,018*	4,959	4,908	*-0,051*	4,302	4,284	*-0,018*	4,116	4,101	*-0,015*	3,931	4,061	*0,13*	
A5	7,947	7,648	*-0,299*	6,892	7,858	*-0,966*	5,833	5,697	*-0,136*	2,492	2,097	*-0,395*	2,707	2,484	*0,133*	4,936	4,449	*-0,487*	4,317	4,295	*-0,022*	4,133	3,664	*-0,469*	4,116	3,023	*-1,093*	
G6	7,439	6,822	*-0,557*				5,561	4,641	*-0,92*	2,319	1,601	*-0,718*	2,482	1,681	*-0,801*	4,819	4,271	*-0,548*	4,079	4,013	*-0,066*	4,254	3,791	*-0,463*	4,125	3,612	*-0,513*	
C7	7,085	6,737	*-0,348*	5,026	4,605	*-0,422*	5,543	5,114	*-0,429*	1,792	1,201	*-0,591*	2,249	1,718	*-0,531*	4,886	4,497	*-0,389*	4,026	4,099	*0,073*	4,252	3,796	*-0,456*	4,011	3,454	*-0,557*	*12,6*
G8	7,754	7,583	*-0,171*				5,789	5,832	*0,043*	2,561	2,411	*-0,451*	2,608	2,649	*0,041*	4,868	4,787	*-0,081*	4,27	4,084	*-0,186*	3,945	3,796	*-0,149*	3,87	3,447	*-0,423*	
C9	7,29	7,194	*-0,096*	5,202	5,221	*0,019*	5,821	5,816	*-0,005*	1,992	1,825	*-0,167*	2,424	2,175	*-0,249*	4,637	4,337	*-0,32*	4,131	4,144	*0,013*	4,175	4,027	*-0,148*	4,084	3,808	*-0,276*	*12,55*
T10	7,349	7,157	*-0,192*	1,506	1,365	*-0,141*	5,909	5,748	*-0,161*	2,021	1,725	*-0,296*	2,451	2,265	*-0,186*	4,766	4,338	*-0,428*	4,074	2,075	*-1,999*	4,036	3,304	*-0,732*	3,982	2,789	*-1,193*	*14.40*
T11	7,314	7,062	*-0,252*	1,633	1,453	*-0,18*	5,608	5,095	*-0,513*	2,026	1,526	*-0,51*	2,356	1,623	*-0,733*	4,801	4,392	*-0,409*	4,043	2,582	*-1,461*	4,016	3,708	*-0,308*	3,991	3,555	*-0,436*	*14.02*
A12	8,823	8,375	*-0,454*	7,314	7,956	*0,642*	6,079	5,749	*-0,33*	2,624	2,542	*-0,082*	2,755	2,809	*0,054*	4,941	4,832	*-0,109*	4,36	4,271	*-0,093*	4,107	4,182	*0,075*	4,033	3,788	*-0,245*	
C13	7,187	7,212	*-0,025*	5,257	5,538	*0,281*	5,539	5,241	*-0,298*	1,756	1,575	*-0,181*	2,18	2,015	*-0,165*	4,672	4,493	*-0,179*	4,059	4,023	*-0,036*	4,162	3,671	*-0,491*	4,012	3,353	*-0,659*	*12.32*
G14	7,791	7,797	*-0,006*				6,026	5,993	*-0,033*	2,257	2,226	*-0,031*	2,485	2,527	*0,042*	4,553	4,596	*0,043*	3,965	4,145	*0,18*	4,078	4,033	*-0,045*	3,954	3,759	*-0,195*	

For each deoxynucleotide proton the value in italics, (column 3) represents the shift calculated from subtracting the duplex value (column 1) from the adduct value (column 2).
Chemical shift differences (italics) were underlined if the difference was greater than 0.25ppm.

Table 4 contains both exchangeable and non-exchangeable proton chemical shifts for adozelesin in the resulting adozelesin-DNA adduct.

H3	H6	CH3	H8A2	H8A	H8B	H1A	H1B	H3'1	H4'1	H6'1	H7'1	H3'2	H4'2	H5'2	H6'2	H7'2
7,99	6,98	2,546	4,495	5,063	5,155	3,689	3,812	7,283	7,805	8,243	7,629	7,651	7,932	7,497	7,373	7,736

However, no estimation could be made for the remaining nucleotides because their peaks seemed to be of nearly the same intensity, probably due to the effects of spin diffusion. This concludes the interpretation of the 2D NOESY spectrum of duplex DNA 5′-$d(C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12}C^{13}G^{14})_2$. All non-exchangeable chemical shifts were entered into Table 3.

3.4 The Adozelesin-5′-d(CGTAACGCGTTACG)$_2$ adduct

After the NMR analysis of the DNA duplex had been completed, it was allowed to react with adozelesin. The resulting adduct was then subjected to further one- and two-dimensional NMR analysis.

3.4.1 The 1D ^{1}H NMR spectrum

Once the DNA duplex which had reacted with adozelesin had been purified by various techniques, a 1D ^{1}H NMR spectrum was taken to see if the drug had bound covalently to the DNA. Each spectrum had exactly the same spectral width, allowing them to be superimposed over one another, so that the differences between the spectra and, hence, the result of the drug reaction would be easily observed

Analysis of the spectrum (Appendix IV Fig. 109) can, once again, commence by analysis of the methyl, signals which are generally strong and clear. The DNA has three methyls which are attributable to thymines T^3, T^{10} and T^{11}, with chemical shifts ranging from δ 1·4 to δ 1·65ppm. Adozelesin alone has only one methyl on its cyclopropapyrroloindole headunit at C7, with a shift slightly downfield at δ 2·07ppm (Walker *et al.*, 1997). After the first incubation of the DNA with drug, these peaks decreased to about half the intensity and multiplied to at least six peaks in the δ 1·25 to δ 1·75ppm range. These represented the three methyls of both the duplex and the covalently modified DNA. There is approximately a 50:50 ratio of each species. A large peak has also appeared at about δ 2·5ppm which is equal in intensity to the DNA methyls, so could be attributed to the covalently bound adozelesin. There does not appear to be a methyl singlet at δ 2·07ppm unless it is obscured within the sugar H2′1, H2′2 protons, suggesting all drug that is present is part of an adduct. The aromatic region probably most clearly exhibits the complexity of the spectrum, when two species are present in the same solution. Seventeen singlets correspond to the DNA duplex purine and pyrimidine H8/H6, plus adenine H2s. This number of resonances again symbolizes the DNA aromatic protons of the drug-DNA adduct.

Also in this region are the eleven aromatic protons from the three ring systems of adozelesin. In all, there are approximately 45 aromatic proton peaks in the space of 1·5ppm.

Further addition of adozelesin was required to take this reaction to completion and yield a single species (Fig. 46). The off scale peak at δ 4·71ppm is due to the solvent HDO. Superimposing this spectrum onto Figure 45, showed that the three large methyl groups were not from DNA duplex but from the adduct, because the positions of the peaks showed movement upfield. A count of the number of aromatic peaks and their shoulders yields about twenty-eight as expected. Because the 1D ^{1}H NMR spectrum of the adozelesin-DNA adduct is so complex, the information gained is limited to this level. Two-dimensional experiments were required for a more detailed analysis.

3.4.2 The 2D ^{1}H NMR DQF-COSY experiment

The initial step in the 2D ^{1}H NMR proton assignment strategy was to solve the DQF-COSY spectrum. COSY peaks result from the coherent transfer of magnetization (scalar coupling) between coupled spins. In analogy with the NOESY spectrum, there is a one-dimensional spectrum which traverses the COSY spectrum along the diagonal but it only contains the proton resonances which are involved in through-bond couplings. The multiplet peaks off the diagonal indicate when spin 'A' is coupled to spin 'X' (Wüthrich, 1986). The phase-sensitive data resolve the multiplet structure into positive and negative signals, so each cross-peak is split into four components, two positive and two negative, separated by J_{AX}. The AX spin-spin coupling therefore produces an 'antiphase square array' of off-diagonal multiplet components. The 400MHz DQF-COSY data was not sufficiently well resolved to enable coupling constant measurements to be made. However, it does contain all the scalar coupling information necessary to yield some definitive starting points for use in the assignment of the 2D NOESY and will affirm all assignments therein, setting them on solid ground. Only homonuclear spin systems were analysed in this 2D ^{1}H DQF-COSY spectrum (Appendix V Fig. 110).

3.4.2.1 The CH5-CH6 region

Analogously to the beginning of the assignment protocol for the 2D NOESY spectrum, the cytosine base H6 to H5, three-bond vicinal coupling, is also a good point at which to commence the assignment procedure in the DQF-COSY. This is because this set of resonances appears in a separate area of the spectrum, free from the confusion of any other proton cross-peaks. Inspection of Appendix V Figure 111, shows that there are four large

antiphase square arrays of cross-peaks propagated from cytosines AX spin system, at δ 5·75, δ 5·54, δ 5·22 and δ 4·60ppm. An upfield shift of at least 0·4ppm in comparison to the most shielded cytosine (C^7), in the duplex NOESY spectrum is observed. It is not possible to assign the peaks to a particular cytosine but it is important to know their positions for reference in the NOESY spectrum of the adduct, because extensive peak overlay or drug induced peak shifting into other regions of the spectrum can make them difficult to identify.

Some other peaks of much weaker intensity than the larger assigned ones are also detectable in this region. These are due to another species in solution, which accounts for approximately ten to fifteen per cent of the population. These peaks are not attributable to unreacted DNA duplex because they do not have the same chemical shifts. Only the most intense peaks of this second species are visible in the DQF-COSY spectrum, which is more sensitive than the NOESY. The cross-peaks of this second species are not only weak in intensity but are too few in number to make any attempt at assignments and, hence, give any conformational information on the binding of adozelesin to DNA in this species. However, there are five hypotheses to what this adozelesin-5′-d(CGTAAGC)·5′-d(GCTTACG) adduct structure could be. One possibility is that only one adozelesin has reacted with the duplex 14-mer, instead of two. Monoalkylation leads to loss of self-complementarity within the adduct, resulting in an area equivalent to unmodified duplex and the other end corresponding to drug-DNA adduct. The cytosine H5 chemical shifts of the minor species (Appendix V Fig. 111) are approximately δ 4·55, δ 5·09, δ 5·74 and δ 5·87 ppm, which is not in concordance with the duplex values of δ 5·02, δ 5·20, δ 5·26 and δ 5·78ppm. The cytosine H6 chemical shift coordinate of the CH5-CH6 cross-peaks also shows degrees of up and downfield shifting. This evidence implies that a monoalkylated adduct is not responsible for the minor species. A second possibility is that one or both of the modified adenine nucleotides in the DNA strand may have depurinated. This process forms the basis of the Maxam-Gilbert sequencing technique (Stryer, 1988). The adozelesin modified adenine (A^{12}) is removed from the DNA strand *via* β-elimination. This involves cleavage of the glycosidic bond generally on heating, leaving the sugar without a base *i.e.* apurinic (Hurley *et al.*, 1984). Although this phenomenon can occur on standing over long periods of time, this adduct was stored refrigerated in its freeze-dried form to discourage this outcome. It is interesting to note that these cross-peaks are more visible in the 600MHz than the 400MHz data. This maybe due to the number of months of sample storage between times of data collection allowing additional depurination, or it could be their intensity corresponds to the increased sensitivity of the experiment. However, the facts that the 400MHz DQF-COSY experiment was run when this sample was relatively fresh (~ 1 week old) and that the same sample produced 600MHz data which were not markedly different even after five months suggest this sample is quite stable. Depurination, it seems, is not the most favourable explanation for the minor species exposed by the 400MHz data. However,

depurination and monoalkylation could be responsible for the other weak peaks in this region, seen only in the 600MHz spectrum.

A more plausible reason for the appearance of the minor species is likely to be the alkylation on an alternative adenine. Seaman *et al.*, (1996) described the bisalkylation of 5′-d($C^1G^2T^3A^4A^{5*}T^6T^7A^{8*}C^9G^{10}$) self-complementary duplex by CPI-I, which has an analogous structure to adozelesin minus the benzofuran subunit (C). Bisadduct formation resulted from the 'head' (CPI) to 'tail' (I) binding of CPI-I at A^5 and A^8. Alkylation first took place at A^8, causing a distortion in the same strand, five base-pairs upstream in the 5′-direction. Modification of the second adenine therefore took place at the position of least distortion. Because this DNA duplex is longer, to accommodate the benzofuran (C) subunit of adozelesin, a distortion of this kind would now be situated around C^7. Adozelesin would be free to bind at both A^{12} and A^5. A^{12} maybe more available for reaction than A^5 to form the major product of an adozelesin 'tail to tail' bis-adduct. Not only is A^{12} furthest away from any possible distortion in the GC region but alkylation is taking place at the preferred 5′$T^{10}T^{11}A^{12*}C^{13}$ consensus sequence (Weiland and Dooley, 1991; Yoon and Lee, 1998). The minor species maybe related to alkylation of A^5, yielding an analogous 'head to tail' bis-adduct to that characterized by Seaman *et al.*, (1996). 5′-d($T^3A^4A^{5*}$) still constitutes a consensus sequence (Weiland and Dooley, 1991), although any distortion around C^7 could discourage binding hence, the production of only a minor species.

Alkylation of the central adenine A^4 would produce a monoadduct. This is because the combined length of the adozelesin subunits A, B and C requires that it span five bases, which would protrude into the second binding site, obscuring a second alkylation. A monoadduct is a possible candidate for the minor species, although alkylation of A^4 could mean that the benzofuran ring of adozelesin protrudes out of the end of the duplex, which is not energetically favourable. The consensus sequence described by Weiland and Dooley, (1991), *i.e.* 5′-d(A/T)(G/C)(A/T)A^* is similar to 5′-d($C^1G^2T^3A^{4*}$) but not identical, suggesting it would not bind with such a high affinity. It is not inconceivable that adozelesin could bind non-covalently within the minor groove to some other site. Gunz and Naegeli (1996) demonstrated that (+)CC-1065 could do this with such affinity that it could not be removed by dialysis or precipitation techniques. However, in the presence of high affinity alkylation sites containing 5′-d(TTA), it is unlikely that the noncovalent binding mode would be able to compete. In the 'tail to tail' bis-adduct (adozelesin modification of A^{12*}), there would be no space in the 14-mer for another adozelesin molecule to non-specifically bind, due to these drugs collectively spanning 10 out of 14 bases.

3.4.2.2 The TH6-TCH_3 region

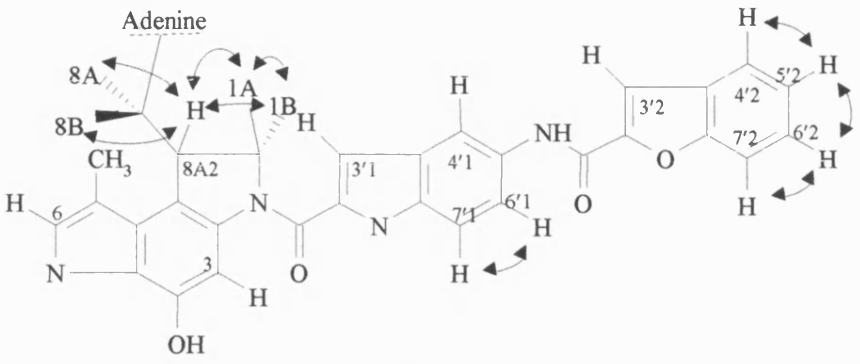
None of the expected TH6-TCH_3 peaks corresponding to thymine A_3X spin system were present in the 2D DQF-COSY spectrum. These would have been a useful tool for separating the TCH_3-TH6 cross-peaks from those of the methyl to aromatic proton of the 5′-flanking nucleotide, in the NOESY spectrum, which can be of similar intensity. Their non-existence could be related to the insensitivity of the 400MHz data to a four-bond (^4J) coupling, where there is quite a long distance between the protons. If the peaks were of very low intensity they would be obscured by the noise.

3.4.2.3 The deoxyribose sugar resonances

The cross-peaks resulting from spin-spin couplings from around the deoxyribose ring, exist in abundance throughout the different regions of the DQF-COSY spectrum (Appendix V Fig. 112, 113, 114). Geminal two bond couplings (^2J) of H2′1→H2′2 and, especially, H5′1→H5′2, produce cross-peaks which are extremely close to the diagonal or possibly even obscured by it, if the chemical shift difference between the protons is very small. There are also a significant number of three bond couplings (^3J) around the ring; H1′→H2′1, H1′→H2′2, H2′1→H3′, H2′2→H3′ and H3′→H4′. However, the DQF-COSY spectrum can only be properly utilized when used in conjunction with the NOESY spectrum, which is lined up alongside it. The cross-peaks close up in the DQF-COSY cannot be assigned to a specific nucleotide, because there are no sequential connectivities between neighbouring nucleotides. However, by aligning this spectrum with the 2D NOESY spectrum, where a starting point such as those identified by cytosine CH5-CH6 COSY couplings has already been established, it is possible to create a cycle of NOESY-COSY connectivities. Sequential connectivity 'walks' between base protons and protons of the 5′-flanking sugar *e.g.* H2′1/H2′2$_{(n)}$→H8/H6$_{(n+1)}$ in the NOESY can have their sugar H2′1/H2′2 assignments checked, by moving perpendicularly to the D2 axis at the chemical shift of each preliminarily assigned H2′1 and H2′2, straight across the COSY spectrum to the diagonal. If there is a cross-peak generated from these co-ordinates, then the chemical shifts can be unequivocally assigned. In the same manner, the H1′, H3′ and H4′ protons were assigned. Once the H2′1 and H2′2 geminal coupling for a sugar was located, it was a simple task to find the vicinal H1′ sugar resonance to which it was coupled and verify it with the NOESY data. From the chemical shift of H1′ the spin-spin coupling to H3′ and from here to H4′ was solved, in agreement with the NOESY spectrum.

3.4.2.4 The aromatic region

There are no DNA protons that give rise to COSY peaks in D1 = 7·0→8·5/D2 = 7·0→8·5 region. However, four strong peaks were detected (Appendix V Fig. 115) and attributed to the aromatic spin-spin couplings within the drug (Fig. 51).



Adozelesin

Figure 51 Spin-Spin couplings and atom labels within adozelesin.

The H6′1→H7′1 vicinal coupling of the indole was assigned to the peak which did not align with any others because it is an isolated spin system, H6′1 δ 8·24ppm and H7′1 δ 7·629ppm. The three remaining peaks are part of a connectivity network, which represented the vicinal couplings around the benzofuran tail. H5′1→H6′2 was assigned to the central peak in the sequence. It was difficult to assign the chemical shift to either proton, without knowing which terminal cross-peak corresponded to H4′2 or H7′2. In adozelesin alone, the chemical shift of H4′2 (δ 7·85ppm) is further downfield than H7′2 (δ 7·67ppm) (Walker *et al.*, 1997). However, when the adozelesin has covalently bound to DNA, this new environment may not induce the same trends in chemical shift. Therefore, using the NOESY spectrum, the NOE cross-peak H3′2→H4′2 was sought and this decisively gave the assignments; H3′2 δ 7·65 and H4′2 δ 7·93ppm, because only H4′2 was close enough to have an NOE connectivity to H3′2. This assignment led to the deduction of the others through their sequential COSY connectivities, *i.e.* H5′2 δ 7·49, H6′2 δ 7·37 and H7′2 δ 7·73ppm.

3.4.3 The 2D ^{1}H NMR NOESY spectrum

Assignment of the 2D 400MHz ^{1}H NMR NOESY spectrum (Appendix VI Fig. 116) began with spectrum produced at 250ms mixing time, because this yielded the clearest, most well-resolved data. The protocol for assigning the proton 'walks' within the DNA of the

adozelesin–5′-d(CGTAAGCGCTTACG)$_2$ adduct is exactly the same as that used for assigning the DNA duplex alone. The complexity of the spectrum, however, requires that the utmost care be taken when assigning peaks, to ensure that drug-drug or DNA-drug NOE conectivities are not incorporated into sequential walks.

3.4.3.1 The cytosine H5 region

The DQF-COSY spectrum has already identified four cytosine CH6-CH5 spin-spin couplings, ranging from δ 4·60 to δ 5·74ppm. The NOESY spectrum (Fig. 52) was then used to distinguish between the resonances, one of which had shifted considerably upfield into the H3′ region. C^{13} is the only cytosine nucleotide adjacent to an adenine, which is typically found to resonate further downfield than C, T or G, as already explained. The ring system of the base A^{12} is covalently modified by the cyclopropapyrroloindole headunit of adozelesin. Therefore, the base proton A^{12}H8 will be in close spatial proximity to the aromatic rings of the CPI subunit of adozelesin and, when A^{12}H8 is positioned perpendicularly to the CPI rings, it will feel the effects of the shielding. All of the adenine protons would be expected to participate in some upfield shifting due to this effect, although the extent of it would much depend on how close they are to the CPI aromatic rings. However, taking into consideration this effect, adenine H8s chemical shift was still expected to be downfield of the other bases. A CH5 peak at δ 5·54ppm correlated directly into a cross-peak whose base proton shift was δ 8·375ppm, consistent with being an adenine. This was given the preliminary assignment of C^{13} and the base proton to A^{12}. The large cross-peak at δ 5·74ppm had two other peaks aligned with it vertical to D1, one of which was slightly more intense than the other H1′→H6/H8 or H6/H8→H1′ peaks in the spectrum, which is consistent with being a terminal. This peak only had an NOE connectivity to one other peak in the direction of D2, so this resulted in the assignment of C^1H1′, C^1H5 and C^1H6. The other resonance along the C^1H6 line was thought to correspond to that of a drug peak, because there were a horizontal line of points traversing the spectrum here. As before, C^7 and C^9 could not be distinguished in this region.

3.4.3.2 The thymine methyl region

Although some T_1 ridging is apparent in this region (Fig. 53), the strong NOE connectivities of the methyls are clearly visible. T^3 was assigned as the methyl which was only connected to its 5′-flanking base G^2, where T^3H6 was assigned upfield to G^2H8: this also agreed with the sequential assignment of G^2H8 relative to C^1H1′. Meanwhile, moving in reverse order, the intense NOE at δ 1·43ppm correlated directly with its 5′-neighbouring H6, which is in vertical alignment (from D1) with the other thymine methyl. An NOE also occurred between

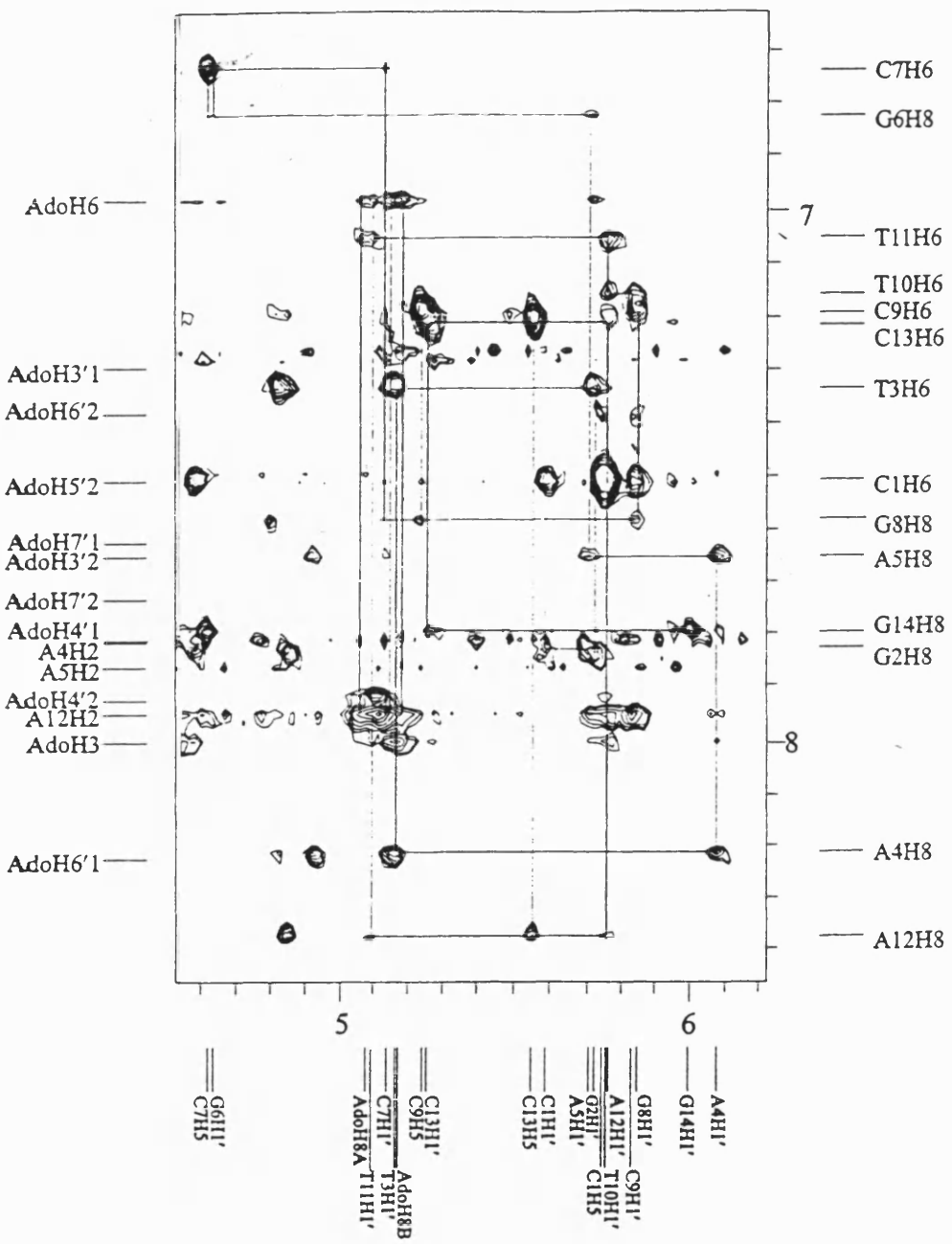


Figure 52 The H6/H8→H1′/H5 region at the two-dimensional 400MHz NOESY spectrum of the 5′d($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$-adozelesin adduct.

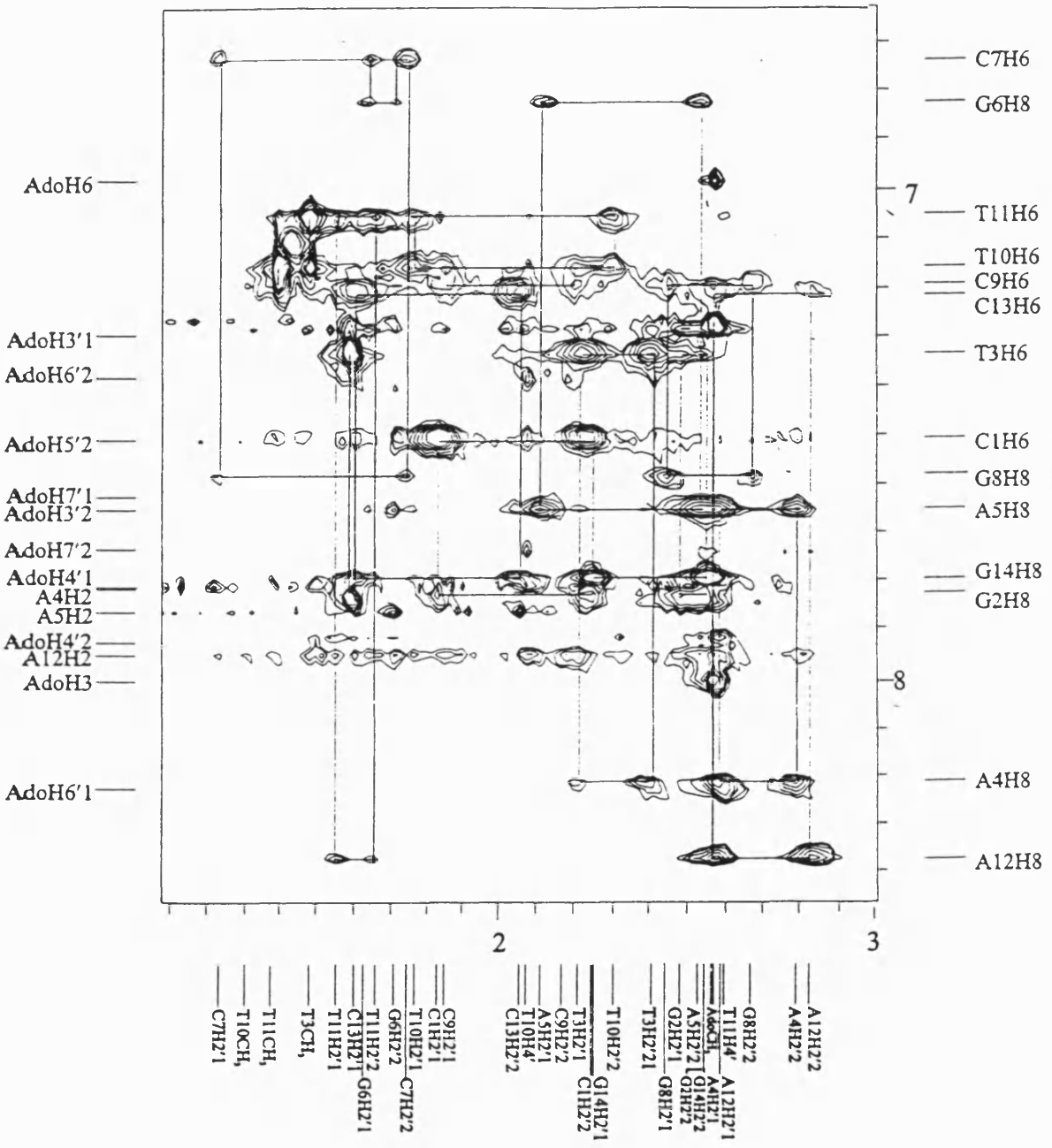


Figure 53 The H6/H8→H2′1/H2′2/CH_3 region of the two-dimensional 400MHz NOESY spectrum, of the 5′d($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$-adozelesin adduct.

the 5′-adjoining base of this second thymine and its methyl, producing a shoulder on the methyl-H6 peak. The former methyl was assigned as $T^{11}CH_3$ and the latter as $T^{10}CH_3$. The realization of the C^9H6 chemical shift (5·22ppm) from $T^{10}CH_3$ led to the automatic assignment of G^8H5 and G^8H8. C^7H5 was assigned by default but G^6H8 could not be assigned from here as the peaks including C^7H5 were relatively weak. Possibly they are too close in chemical shift to the D_2O reference point and to the point of the strong water suppression, yielding a slightly negative baseline dip, pulling adjacent peaks into it.

At this point of the NOESY spectral analysis, preliminary base proton assignments have been achieved for 5′-d($C^1G^2T^3$---$C^7G^8C^9T^{10}T^{11}A^{12}C^{13}$---)$_2$3′. Blank spaces represent the nucleotides which have not been located. Even though adozelesin has now bound to the DNA, one set of NOE connectivities only have been obtained corresponding to one strand of DNA, as was the case with the DNA duplex. To retain self-complementarity an adozelesin must be bound to A^{12} of each strand, so that each end of the DNA adduct is identical. This yields a drug:DNA binding ratio of 2:1 for this complex.

3.4.3.3 The H2′1/H2′2→H6/H8 region

The H2′1 and H2′2 to C^1H6 strong terminal NOEs are easily pinpointed in the NOESY spectrum (Fig. 53), providing a good starting location for the sequential walk between deoxyribose protons and bases, in the 5′→3′ direction. These intense intranucleotide connectivities C^1H2′1/H2′2→C1H6 display correlations into their 3′-flanking base G^2H8 in the horizontal direction towards D2. This internucleotide connectivity is much weaker, due to the longer distance between protons of adjacent nucleotides. The pattern of two strong and two weak NOEs for each nucleotide only continues as far as A^5, where unexpectedly all peaks are of equally very strong intensity. Cross-peaks corresponding to A^5H2′1/A^5H2′2→G^6H8 are weaker than the intranucleotide NOEs of A^5, as anticipated. However, the G^6H8→G^6H2′1/H2′2 NOEs are exceptionally weak, the cross-peaks being less than a third of the size of the internucleotide ones of G^6 to A^5. The chemical shift of G^6H8 is also positioned particularly far upfield for a guanine base at δ 6·82ppm, when normally it is somewhere downfield of δ 7·3ppm. Through space connectivities related to C^7 are also weak, although its intranucleotide NOEs are its most intense peaks, with G^6H2′1/H2′2→C^7H6 producing slightly more intense peaks than G^6H2′1/H2′2→G^6H8. The reversal of inter- and intra-nucleotide peak intensities related to G^6H8 seems to indicate an inversion of interproton distances, with G^6H8 being closer to its 5′-flanking nucleotides (A^5) sugar protons than its own. This may indicate that more than a flip in sugar conformation *i.e.* from 2′C endo to 3′endo (β to A-DNA) is taking place. These results seem to be more consistent with a distortion of the nucleotide in the β-helix, because the base-

sugar 'walk' is still present. NOEs proceeding from G^8 to C^9 show a steady increase in peak intensities, with the return of the normal β-form pattern of weaker intranucleotide and stronger internucleotide connectivities, proceeding sequentially along the DNA strand.

3.4.3.4 The H1′→H6/H8 region

Sequential assignment of the deoxyribose H1′ protons (Fig. 52) was carried out as previously described for the duplex DNA. The C^1H1′→C^1H6 resonance at co-ordinates δ 5·58 and δ 7·51ppm, respectively, was the starting point. It is interesting to note, that by analogy to the description of exceedingly weak NOEs corresponding to G^6H8→H2′1/H2′2, the peaks in the H1′ 'walk' beginning with A^5H8→A^5H1′ through to G^8H8→G^8H1′ are also exceptionally weak, suggesting some perturbation of the β-helix on adozelesin binding. The sequential 'walk' then continues uninterrupted through to G^{14}H8→G^{14}H1′.

3.4.3.5 The H1′→H2′1/H2′2 region

Inspection of this zone of the spectrum (Appendix VI Fig. 117, 118) verified that each assignment was correct when the deoxyribose H1′, H2′1 and H2′2 chemical shifts were mapped together, and the resulting intersection was at the centre of a cross-peak. The most intense NOE of each H1′→H2′1, H1′→H2′2 pair was assigned to H2′2, due to the much shorter distance H1′→H2′2. All H2′2 resonances were generally downfield of H2′1, because of the deshielding ring currents effects of the purines and pyrimidines. However, A^4, A^{12} and G^{14} were an exception to this trend, with their H2′2 proton chemical shift upfield of H2′1. The swapping of H2′1 and H2′2 resonance positions is easily explained for G^{14}, because this is a common occurrence for 3′-terminal nucleotides which do not have a negatively charged 3′-phosphate. This phosphate group normally has a polarizing effect on H2′1 and H2′2, sending the former upfield and the latter downfield. A^4H2′2 and A^{12}H2′2, however, must be experiencing some shielding effect by the adozelesin, resulting in the upfield shift of H2′1.

Referring back to the H6/H8→H2′1/H2′2 section of the spectrum after distinguishing between H2′1 and H2′2, an analysis was made of whether the intranucleotide cross-peak H6/H8→H2′1 was more or less intense than that to H2′2. In eight out of fourteen nucleotides the H2′1→H6/H8 did appear to be more intense, leading to the crude assumption that their glycosidic torsion angles were in the anti conformation (Feigon *et al.*, 1983b), typical of β-form DNA. No measurements were able to be made as the data was not of good enough quality. Of

the remaining nucleotides, G^2, C^9, T^{11} and C^{13}, it was impossible to tell which was more intense, partly due to the presence of the t_1 ridging. C^7 and A^5 however, had H2′2→H6/H8 NOE connectivities which were more intense. By default, this could be taken as an indication of syn conformation of the glycosidic torsion angle. This is generally observed in Z-DNA but the evidence so far strongly indicates β-form DNA, as all sequential walks are present along the DNA strand, even if the intensities of some nucleotides are much weaker than expected, implying some helical perturbation. It is more likely that the distance between the H2′1/H2′2→H6/H8 is a product of both a change in the glycosidic torsion angle and the sugar conformation *i.e.* C3′-endo (A-DNA) or C2′-endo (β-DNA). Accurate measurement of distances and *J* values from the sugar ring protons to the base aromatic protons is required to determine sugar conformation. C2′-endo geometry is denoted by strong NOE cross peaks H1′→H2′1, H1′→H2′2 and H2′1→H3′, whereas H2′2 may or may not display its NOE to H3′ because it is so weak (Kintanar *et al.*, 1987; Wijmenga *et al.*, 1993). An accurate analysis of the H3′→H2′1/H2′2 region was not possible with 400MHz data, because the data was partly obscured by T_1 ridging and the effects of base-line roll.

3.4.3.6 The H3′, H4′, H5′1, H5′2 protons

The H3′→H6/H8 (Appendix VI Fig. 119) and fragments of the H4′→H6/H8 were assigned using previously described procedures. Due to the H4′ walk being quite weak in the H6/H8 region, assignments were made mainly from H1′→H4′ and H3′→H4′, where the interproton distances were shorter and the effects of spin diffusion were more pronounced. Both H5′1 and H5′2 resonances were assigned *via* their intranucleotide connectivities to H3′and confirmed by those to H4′ (Appendix VI Fig. 120, 121). H5′2 was assigned upfield of H5′1 in accordance with Remin and Shugar's (1972) rule.

3.4.3.7 The Adenine H2 protons

A preliminary identification of the whereabouts of these proton signals was obtained by examination of the 1D spectrum (Fig. 46), where three tall, sharp peaks were noted two thirds of the way across the aromatic region. Correlation with the 2D NOESY (Appendix VI), demonstrated AH2 type peak smears traversing the spectrum at these points. Assigning each AH2 to A^4, A^5 and A^{12} was achieved by observation of their NOEs to deoxyribose H1′ protons.

This concluded the assignment of the non-exchangeable sugar and base protons of DNA. All proton chemical shifts are given in Table 3. The chemical shift difference was

calculated for each deoxyribonucleotide by subtracting the duplex value from the adduct value. The change in chemical shift was then represented graphically (Fig. 54) to illustrate which protons experienced the greatest shifts on the addition of adozelesin to DNA. The sequential connectivity network, with its distinct pattern of peak intensities along the DNA strand, indicated that β-helical structure with Watson-Crick base pairing had been preserved.

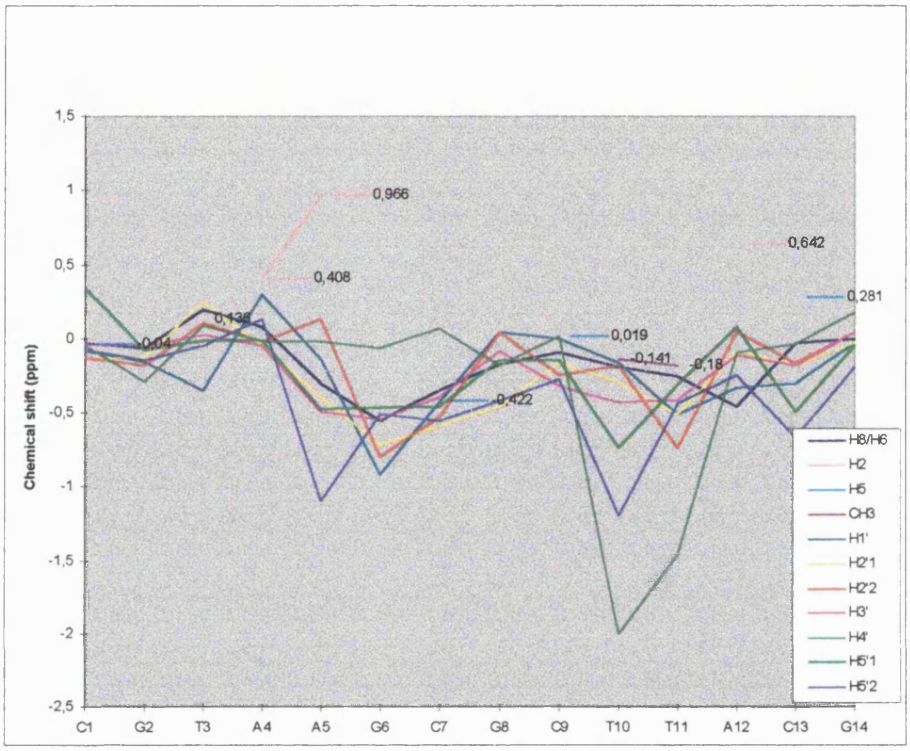


Figure 54 Graphical representation of the chemical shift changes that occur for each proton in DNA, as a result of the covalent binding of adozelesin, to form an adduct.

3.4.4 ^{1}H NMR assignments of the adozelesin non-exchangable protons

Adozelesin was divided into its three constituent subunits; a cyclopropapyrroloindole (CPI) headunit (A), an indole (B) and a benzofuran tail (C), to facilitate assignment of the non-exchangeable protons, The aromatic protons of the benzofuran sub-unit (C), have already been assigned in the DQF-COSY spectrum, due to the large ortho couplings around the aromatic ring systems. The through-space connectivities between these protons H4′2, H5′2, H6′2 and H7′2 can also be seen in the NOESY, with the additional NOE between H3′2 and H4′2 (Appendix V Fig. 115).

The indole subunit of adozelesin exhibits a three spin system (Walker *et al.*, 1997) where H7′1 and H6′1 are strongly ortho coupled to each other, as identified previously in the DQF-COSY. The weaker coupling to H4′1 was only identifiable as an NOE, and similarly from H4′1 to H3′1.

Within the CPI subunit which is covalently bound to A^{12} *via* N3, a series of cross-peaks occur in the NOESY data (Appendix VI Fig. 123) between H8A, H8B, H8A2, H1A and H1B. It is immediately obvious from analysis of these protons that, in order to bind to the adenine, the cyclopropyl ring system has opened. The cyclopropyl H have shifted downfield from H8A δ 1·47ppm and H8B δ 2·02ppm (Walker *et al.*, 1997) to H8A δ 5·06ppm and H8B δ 5·15ppm, into the region of the NOESY occupied by methylene protons (Scahill *et al.*, 1990). The methine proton H8A2 was easily assigned, as the multiplet at δ 4·49ppm, resulting from the surrounding ortho coupled protons. The H1A and H1B doublets, with chemical shifts of δ 3·69ppm and δ 3·81ppm respectively, were both responsible for NOE cross-peaks to the nearby H3′1 on the indole subunit. The methylene proton H1A is nearer to H3′1 and was, hence, assigned upfield to H1B, relative to cross-peak size. The H8A and H8B methylene proton doublets (at δ 5·06 and δ 5·15 ppm respectively) gave NOEs with greater intensity to A^{12}H2, than H1A and H1B which were much weaker. H8B was assigned downfield of H8A because the cross-peak to H6 was of greater intensity and, hence, nearer to this aromatic proton than H8A. Some drug-drug proton NOEs were also noted between H8A/H8B→H6, H3 and CH_3 The aromatic H6 proton was identified due to its particularly intense cross-peak with the adjacent CH_3, in the H2′1/H2′2 region of the spectrum. The last unidentified aromatic signal was assigned as H3 based on NOE cross-connectivities to adenine (A^{12})H2. The complete proton assignments for adozelesin are given in Table 4.

3.4.5 Assignment of NOE connectivities between DNA and adozelesin

Once complete proton assignments of DNA and adozelesin had been achieved, the two-dimensional NOESY spectrum was scrutinized for NOE connectivities between 5′d($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12}C^{13}G^{14}$)3′ and the CPI(A), indole (B) and benzofuran (C) subunits of adozelesin. This involved searching methodically for all sugar resonances which intersected those of adozelesin. In all, 35 DNA/adozelesin connectivities were found, confirming the location of the three adozelesin subunits within the minor groove. These connectivities are illustrated in Figure 55. Additionally there were 24 intradrug NOEs from proton connectivities within adozelesin, plus 285 NOEs relating to DNA base-base, base-deoxyribose and deoxyribose-deoxyribose proton connectivities. This led to a grand total of

344 NOE cross-peaks for distance-range constraint generation, followed by their application to the DNA-adozelesin adduct molecular model.

3.4.6 600MHz DQF-COSY, NOESY, TOCSY and ROESY data

After assignment of the 400MHz Jeol 2D ^{1}H NOESY and DQF-COSY data had been completed and a refined molecular model of the adozelesin-DNA adduct had been attained, 600MHz Varian 2D ^{1}H DQF-COSY, NOESY, TOCSY and ROESY data were made available, courtesy of Dr. George Gray at Varian, Palo Alto. Comparison of proton assignments made from the 400MHz data, where peaks were often somewhat weak, showed a perfect correlation with the clear, well-resolved 600MHz spectra, placing all NOE connectivities and proton assignments on solid ground.

The 600MHz DQF-COSY and NOESY spectra are displayed in Appendix VII and VIII respectively. The enhanced sensitivity of the 600MHz NMR experiment allowed identification of another 33 adozelesin-DNA cross-peaks. Some of these NOE connectivities were also present in the 400MHz data. However, they were not initially included in distance range constraint generation because these cross-peaks were so weak that they could not be distinguished with confidence from interference by noise. Only computer printouts were available for these 600MHz data sets, so no quantitative information on these NOE connectivities was extracted. These extra NOE adozelesin-DNA interproton connectivities verified the drug orientation and positioning of each subunit, relative to each nucleotide in the DNA strand. The 35 distance range restraints from the 400MHz data pushed and pulled adozelesin into position within the minor groove during molecular mechanics and dynamics calculations. Measurements of the 33 extra NOE interproton distances on the computer generated model, produced distances of generally around 5Å in concordance with weak peaks in the 400MHz data being at the distance limits of NOE detection.

TOCSY data (Appendix IX) provided information on multi-step scalar couplings around for example, the aromatic protons of the benzofuran ring of adozelesin, or from proton to proton around the sugar ring H1′→H2′1/H2′2→H3′. ROESY data (Appendix X) only displayed the most intense cross peaks, with through space NOEs resulting from spin-diffusion being much weaker. ROESY data can provide information on conformational exchange within the DNA helix. Seaman and Hurley (1996) examined the covalent modification of 5′-d(TAATTA)$_2$ by the interstrand cross-linker bizelesin.

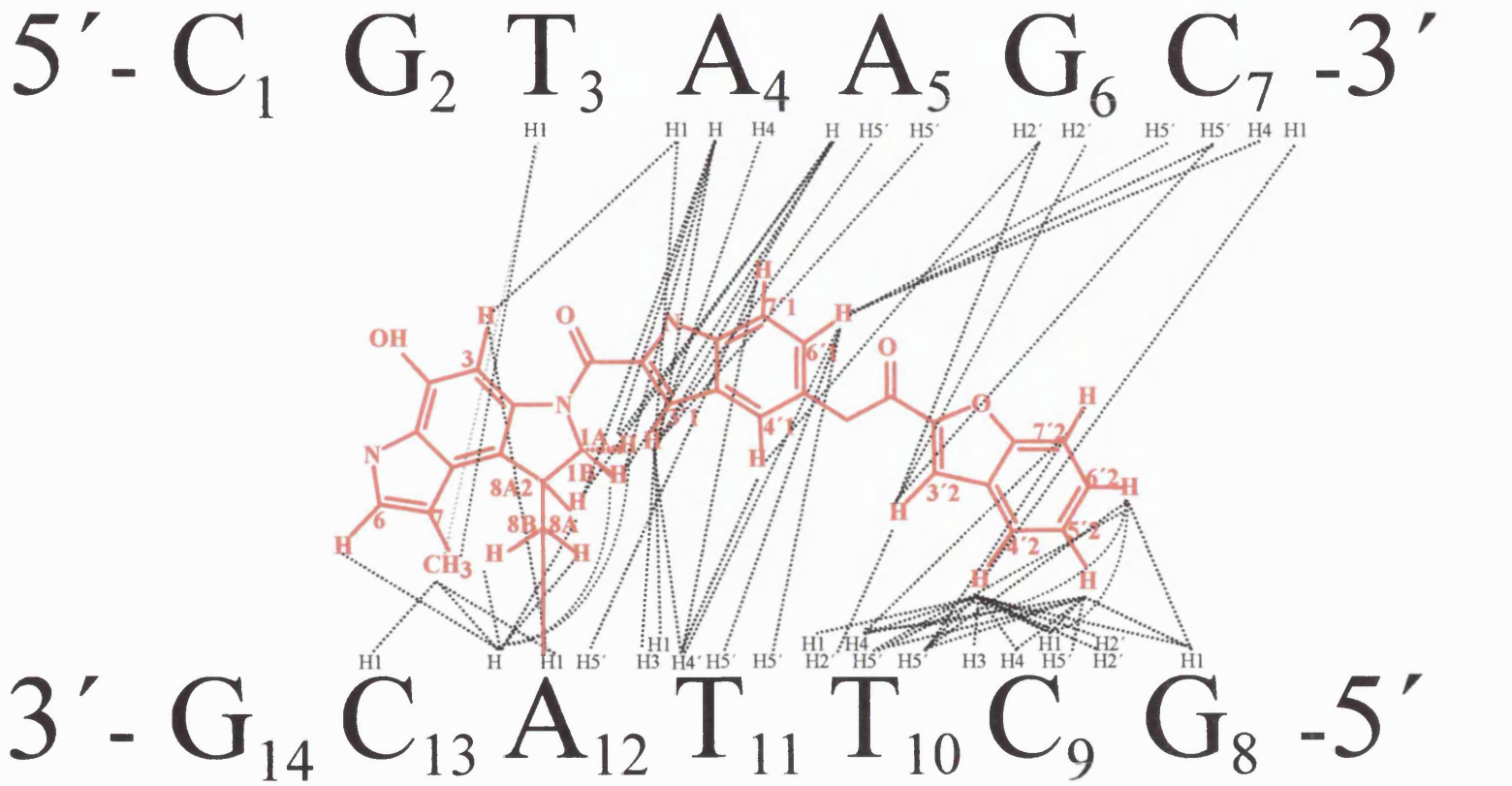


Figure 55 NOE connectivities between DNA and adozelesin.

The ROESY spectrum exhibited strong and weak negative cross-peaks which corresponded to molecular movement between a major and minor isomer. The major isomer contained a central AT step in which the adenines were in syn-orientation, leading to unusual Hoogstein instead of Watson-Crick base pairing. In the minor isomer, this same AT step was in an open conformation, with no hydrogen-bonding between base pairs. The 600MHz ROESY data corresponding to the 5′-d(CGTAAGCGCTTA*CG)$_2$-adozelesin adduct did not reveal any evidence for molecular movement of this nature. This indicated all base-pairs were Watson-Crick style, in concordance with β-helical DNA.

3.4.7 DNA and Adozelesin Exchangable Proton Assignments

Two-dimensional NOE (H_2O) experiments were conducted (also courtesy of Dr. George Gray, at Varian, Palo Alto) and the NH and OH signals were assigned based on their connectivity networks into the aromatic region and imino protons of neighbouring bases. All exchangable signals were identified within the duplex adduct, except for those associated with the terminal bases. The imino signal of the covalently modified adenine (A^{12}) was significantly shifted upfield, as previously reported for (+)CC-1065 (Lin and Hurley, 1990), and bizelesin (Seaman and Hurley, 1993; Thompson and Hurley, 1995; Thompson *et al.*, 1995), partially confirming the site of attachment. The CPI sub-unit (A) and indole sub-unit (B) NH amine protons were identified by connectivities into the neighbouring aromatic protons. The CPI phenol proton was identified on the basis of connectivity into the neighbouring aromatic H3. The chemical shift of the phenolic proton at δ 13·52ppm is consistent with the δ 13·62ppm observed within the previously reported (+)CC-1065 adduct (Lin *et al.*, 1991) and δ 13·05 and δ 13·80ppm within the A-tract bizelesin adduct (Thompson and Hurley, 1995). However, it contrasts with δ 11·06 and δ 11·10ppm observed within the bizelesin 7-base-pair cross-linked adduct (Thompson *et al.*, 1995). Studies using ^{1}H NMR coupled with ^{17}O-labelled water and phosphates in the (+)CC-1065 adduct confirmed that the phenolic proton in the (+)CC-1065 adduct was hydrogen-bonded (*via* an ordered water molecule) to the phosphate backbone (Lin *et al.*, 1991). Based on the similarity of chemical shifts observed, it would appear that a similar hydrogen bond is formed within this adduct. Signals for the amide NH proton between the indole (B) and benzofuran unit (C) could not be identified in either one or two dimensional ^{1}H-NMR data. A summary of the exchangable DNA proton assignments for the DNA-adozelesin adduct are given in Tables 3 and 4.

3.5 Production of the refined adozelesin-DNA adduct structure by restrained molecular mechanics and dynamics calculations

Following complete assignment of all non-exchangable protons within the adozelesin-DNA adduct, distance-range constraints were generated from the 400 MHz NOESY spectrum. These were applied to a computer molecular model using molecular mechanics and dynamics calculations. The result was a visual representation of the adozelesin-DNA adduct, as defined by the data extracted from the NOESY spectrum.

3.5.1 The importance of water in producing a refined adduct structure which accurately represents that found in a biological system.

Nucleic acids *in vivo* first in the aqueous, buffered environment contained within the nucleus of the cell. Their macromolecular tertiary structure is maintained by the attractive and repulsive forces between charged atoms, of which they consist. β-Helical structure of canonical DNA is sustained by an equilibrium between (Westhof *et al.*, 1995):

1) Electrostatic forces originating from the negatively charged phosphates.
2) Hydrophobic interactions between the aromatic rings of the bases, promote orderly base-pair stacking.
3) Energy of the sugar-phosphate backbone conformation.

Noncovalent bonds within the helix *e.g.* electrostatic, hydrogen and van der Waals forces, are profoundly influenced by the presence of water in biological systems. Water molecules are highly polar with a capacity for hydrogen-bonding, making it a perfect solvent for solvating charged oligonucleotides. Due to the high degree of organisation within water itself and around ions, water can reduce the strength of electrostatic interactions by a factor of 80 (dielectric constant of water), relative to those experienced by DNA alone in a vacuum (Stryer, 1988). Therefore, in an attempt to imitate accurately the structure and movement of the adozelesin-DNA adduct within a biological system, it is mandatory that water be included as an integral part of the structure. The sample itself underwent NMR spectroscopic analysis in aqueous buffered conditions. Na^+ counterions were directly co-ordinated to the phosphate of the backbone, to portray the buffer stabilizing DNA. These were partially hydrated by Molecular Silverware systematic packing of waters around the adduct structure with associated sodiums.

Water contributes to the stability of the three-dimensional conformation of DNA (Westhof *et al.*, 1995). The phosphates constituting the sugar-phosphate backbone are

orientated to maximize favourable interactions with the solvent. Water screens the anionic charge of the phosphate oxygens. The esterified oxygens O3′ and O5′ of the backbone are hydrated to a lesser extent. All hydrophilic polar atoms, *e.g.* deoxyribose ring oxygen, are orientated to hydrogen-bond with water. Hydrophobic bonding between aromatic rings and hydrogen-bonds constituting Watson-Crick base pairing-yield a water-free hydrophobic core. Here, the hydrogen-bonding between polar atoms is at maximum strength, holding the two oligonucleotide strands together in a β-helix. The solvent structure is influenced by polar atoms for example, DNA phosphate oxygens can result in cones of hydration, water columns, bridges, filaments and spines of hydration (Westhof *et al.*, 1995).

There are two methods of simulating the solvent. Implicit treatment of the solvent, describes water as a macroscopic entity, using a distance-dependent dielectric constant (Σ) to reflect electrostatic interactions. Although the calculation is much shorter and simpler, implicit treatment of the solvent does not truly represent the biological system (Westhof *et al.*,1995). Therefore, the adozelesin-DNA adduct was solvated by 4000 Gasteiger-Hückel charged water molecules into a droplet. This microscopic description of a solvated biological system is known as explicit treatment of the solvent. All molecules within the droplet are subjected to molecular mechanics and dynamics calculations. Modelling this huge solvated system gives rise to many problems. A long thermalization and equilibration protocol is required to minimize solvent-solute collision induced conformational changes, which could disturb the natural oscillation between the low energy structures of the adducts during dynamics.

In aquo calculations are also computationally expensive in terms of time (5 days for 10,000 steps convergence gradient minimization, 5 days for a 50ps dynamics run). Macromolecular motions can vary greatly in amplitude from hundredths of an ångström to tens of ångströms and over different lengths in time (from femtoseconds to picoseconds, seconds or even longer) (Karplus and Petsko, 1990). Molecular dynamics simulations give an important insight into molecular structure and motion around a global energy minimum on the potential energy surface. A nanosecond is thought to be required for accurate representation of molecular dynamics; however, this was beyond the computational limits of the Silicon Graphics Indy RG 3000 workstation. The solute-solvent non-bond cut-off value was set to 8Å, in an attempt to curb and simplify the calculation, although this would not be present in an ideal system.

3.6 SYBYL software suite

Much of the research cited in the literature uses AMBER or INSIGHT II and DISCOVER to structurally refine X-ray crystallographic or NMR data. Other software

packages are also incorporated, such as MOPAC for charge calculations, MARDIGRAS for constraint generation and MIDAS for display purposes. However, the SYBYL software suite was utilized during this research project for reasons of availability. It is a general software package with interfaces for NMR processing (TRIAD 6·3), a Biopolymer module for protein, DNA and RNA macromolecule sketching, and the SYBYL interface for molecular dynamics and mechanics calculations. All structure refinement calculations in this study implemented the Tripos force field. The SYBYL 6·2 Theory manual (1995) states that the previous version of SYBYL 5·4 has validated its force field to provide energy values identical to those computed with AMBER version 3·0A.

3.7 The refined molecular models of the adozelesin-5′-d(CGTAAGC)·5′-d(GCTTA*CG) adduct.

Molecular mechanics and dynamics calculations on the docked adozelesin within the 5′-d($C^1G^2T^3A^4A^5G^6C^7$)·5′-d($G^8C^9T^{10}T^{11}A^{12}C^{13}G^{14}$) duplex, which incorporated all distance-range constraints, were performed *in vacuo* and *in aquo* (solvated as a droplet). A complete list of the distance-range constraints is given in the peak pick spreadsheet in Appendix XI, which was generated in conjunction with the master spreadsheet (Appendix XII). Columns 6 and 7 represent the lower and upper bounds (in Å) for the distance-range constraints, which were applied to the adozelesin-DNA model within SYBYL. Nuclear spin-spin coupling, especially NOEs and chemical shift are both affected by internal molecular motion. When interconversion of local conformations is rapid on the NMR timescale, then average values of these parameters are obtained (Karplus and Petsko, 1990). It is therefore justified to calculate the average structure from the 50ps dynamics trajectory, where the generated structures correspond to low energy conformations around a global minima. Averaging of the dynamics data produced a structurally refined molecular model of the adozelesin-DNA adduct to best-fit the average conformation defined by the NMR data. Average structures from the last 50ps of each dynamics trajectory for both systems *i.e.* with and without water, correlated very closely to one another, as well as with the two-dimensional COSY and NOESY data.

Analysis of each trajectory demonstrated how the binding of adozelesin within the minor groove of DNA, acted like a third strand encased by each DNA oligomer. Adozelesin binding to the minor groove stiffened the DNA helix considerably, limiting its mobility. This DNA stiffening effect has been previously noted in two-dimensional gel electrophoresis studies on the circularization efficiency of (+)CC-1065 modified DNA (Sun and Hurley, 1992b). (+)CC-1065-DNA adducts exhibited only low levels of circularization whilst (+)ABC, an analogous structure to adozelesin except for the C ring being an indole instead of a benzofuran,

showed 80 per cent of the population was circularized. (+)ABC has similar topology to adozelesin, making it very probable that adozelesin would produce some stiffening of the DNA helix but not to the extent of that observed with (+)CC-1065 which results in winding of the helix. Winding of DNA was not observed with the compound (+)ABC (Fig. 17) which implies winding is induced by the presence of the ethylene bridge (Lee *et al.*, 1991). It is the ethylene bridge (+)CC-1065 which is thought to be responsible for the winding and associated delayed hepatotoxicity seen with this drug (McGovren *et al.*, 1984; Warpehoski and Bradford, 1988). Adozelesin binding visibly stiffens the helix in the dynamics trajectory, with movement and bending of the helix being limited in the area of the bound drug.

Figure 56 shows the *in vacuo* computer model of the adozelesin-5′-d(CGTAAGC)·5′-d(GCTTA*CG) DNA duplex adduct. Adozelesin is in yellow, the 5′-C^1 strand is in magenta and the 5′G^8 modified strand is blue. Watson-Crick base pairing has obviously been maintained, in agreement with the information from the NOESY spectrum which indicated β-helical structure has been maintained. Even though this sequence is only seven base-pairs in length, which is not a full helical turn of 10·5 ± 1bp (Rhodes and Klug, 1981), it is plain that the DNA has bent slightly to produce a straighter minor groove in which the adozelesin sits. Incorporation of the whole of adozelesin into the minor groove requires a small local bend of the DNA helix axis, so that the third subunit is not protruding out of the groove into space. Powers and Gorenstein, (1990) studied the (+)CPI-$CDPI_2$-5′-d$(CGCTTAAGCG)_2$ adduct by 2D NMR and refined molecular modelling. CPI-$CDPI_2$ was also found to distort the helix on binding to the minor groove, yielding a bent structure. The C^{13}-G^2 base pair is able to accommodate the end of the CPI headunit with the $AdoCH_3$ directed towards G^2.

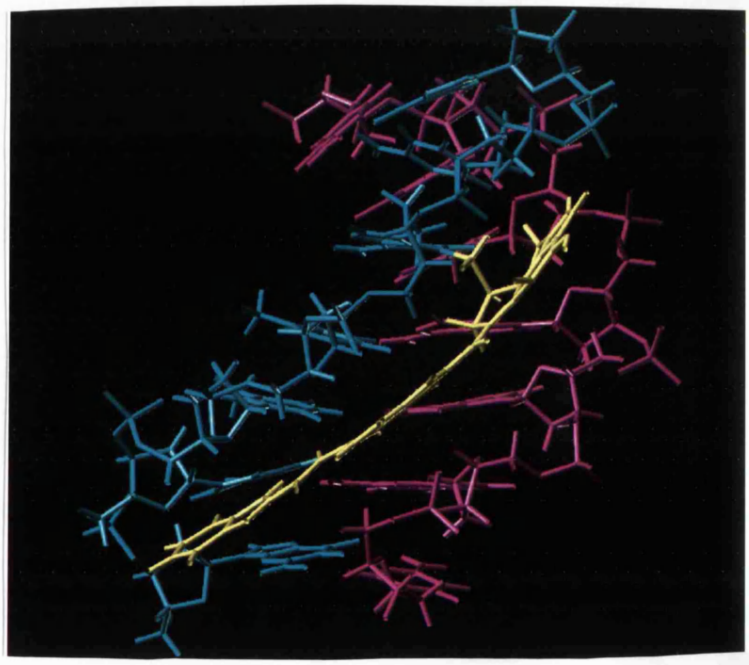


Figure 56 The refined molecular model of the adozelesin-5′-d(CGTAAGC)·5′-d(GCTTACG) DNA adduct, generated *in vacuo*.

A GC base pair may be preferred on the 3′-side of the binding site, because it provides a wider environment for the incorporation of the bulky methyl than the narrower mixed AT tract.

The refined molecular model of this adozelesin-DNA adduct resulting from the *in aquo* calculation is displayed as a 3D stereoview in Figure 57. The colour scheme is identical to that described for the *in vacuo* model. The water molecules which surrounded the adduct in its solvated form, are not shown so as to provide a clear view. The features mimic those observed in the *in vacuo* model. It is clear that adozelesin squeezes itself into the minor groove with a high degree of complementarity, with only its hydrophilic surface exposed to the surrounding solvent. Analogous to (+)CC-1065 (Chidester *et al.*, 1981) and in common with many other classical minor groove binders such as the oligopeptide antibiotics *e.g.* netropsin and distamycin (Berman *et al.*, 1979; Gurskaya *et al.*, 1979), the bis-benzimidazoles *e.g.* Hoechst 33258 (Searle and Embrey, 1990) and the bisamidines *e.g.* berenil (Pearl *et al.*, 1987), adozelesin is crescent shaped and binds isohelically within the minor groove, following the right handed twist of the β-DNA helix.

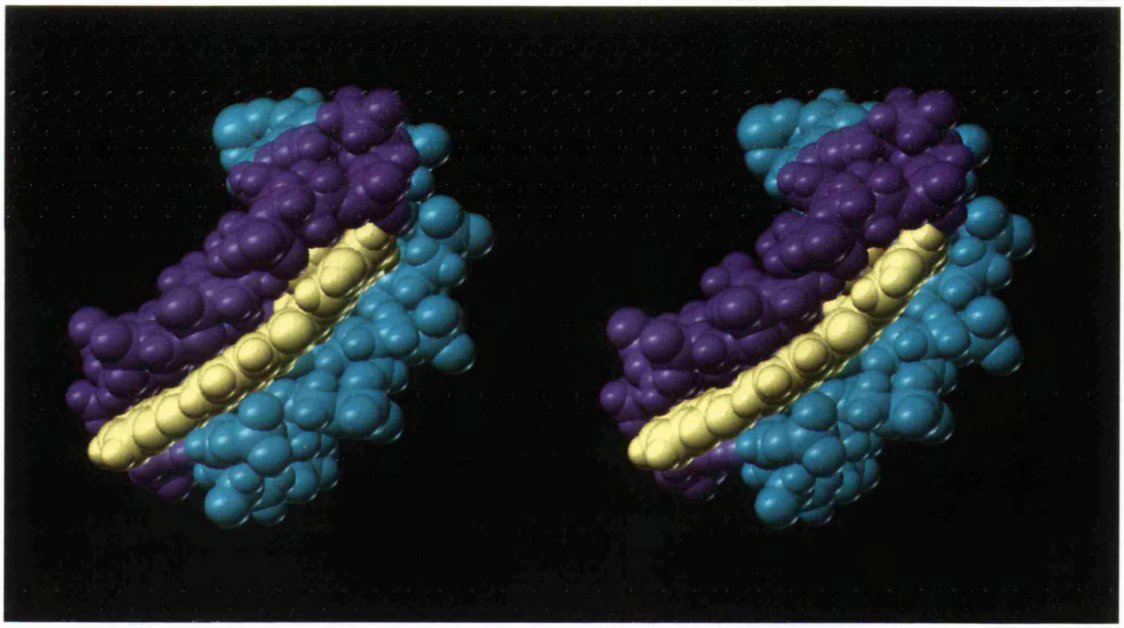


Figure 57 Stereoview of the refined molecular model of the adozelesin-5′-d(CGTAAGC)·5′-d(GCTTA*CG) DNA adduct, generated *in aquo*.

3.7.1 Adozelesin covalently binds to the minor groove of DNA *via* the N3 of adenine A^{12}

The presence of only one complete set of cross-connectivity walks between deoxyribose-deoxyribose and deoxyribose-base protons indicates that there is only one major species of adozelesin-DNA adduct formed. This single proton-proton walk also demonstrates that the palindromic DNA duplex 5′-d($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12}C^{13}G^{14}$)$_2$ has retained

self-complementarity on the binding of two adozelesin molecules to form a bisadduct. Each adozelesin has to bind to an equivalent adenine (A^{12}), in the same orientation (tail extending in the 5′-direction) with the same molecular conformation. Hence, the NOE connectivities between adozelesin and DNA are equivalent for each binding site. Taking into account the symmetry of this adduct, both drugs must either bind to A^5 or to A^{12}. Covalent modification of A^5 would yield a 'head to head' style adduct, where the CPI subunit (A) is the head. Whereas alkylation of A^{12} would produce an adduct where the 'tails' pointed towards the centre of the duplex. Alkylation was predicted to take place at A^{12} because 5′-d($T^{10}T^{11}A^{12*}C^{13}G^{14}$) is the preferred consensus sequence for adozelesin (Weiland and Dooley, 1991; Yoon and Lee, 1998).

A large body of NMR evidence collected from both 400MHz and 600MHz two-dimensional NOESY data sets was available to testify to the covalent modification of the N3 of adenine (A^{12}) by adozelesin. The mechanism of adozelesin binding (Fig. 58) is equivalent to that described for (+)CC-1065 (Scahill *et al.*, 1990).

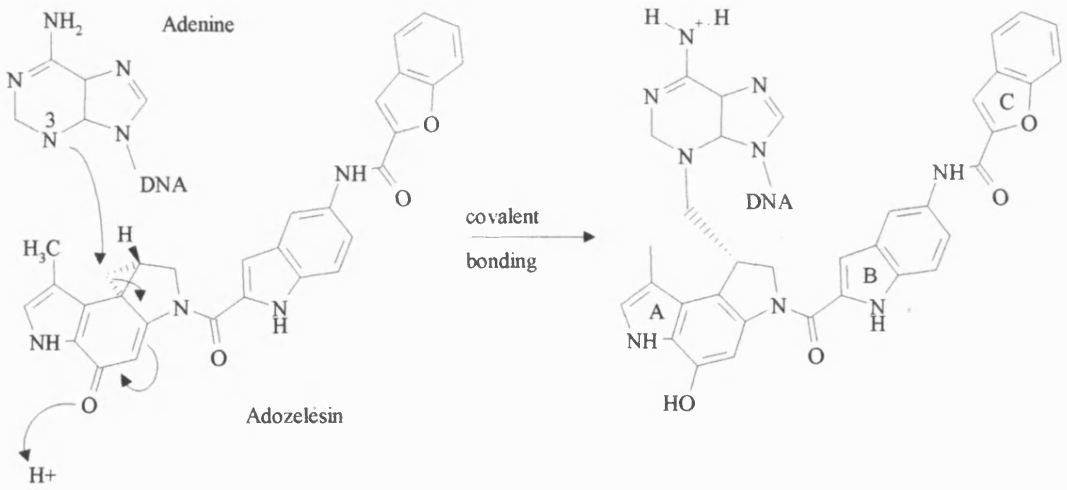


Figure 58 The mechanism of the alkylation reaction between the N3 of adenine and adozelesin.

Alkylation of DNA occurs *via* a concerted reaction in which the cyclopropyl ring opens and the indole quinone of the (A) subunit is reduced to an indole phenol. Downfield shifts of the protons of the cyclopropyl ring *i.e.* H8A at δ 1·47ppm and H8B at δ 2·02ppm (Walker *et al.*, 1997) into the methylene region (δ 5·06 and δ 5·15ppm, respectively) demonstrated that covalent bond formation with adenine was *via* the C8 methylene bridge. Boger and Mésini (1995) demonstrated that A-N3 alkylation took place at the least substituted cyclopropane carbon, which in adozelesin is C8. The adenine A^{12}H2 proton is positioned in the centre of the base of the minor groove, making it an excellent marker to reference the location of adozelesin to. A^{12}H2 shows NOE connectivities to the protons of the CPI headunit which are closest to the

bottom of the minor groove. Adozelesin H3 is situated on the opposite side of the ring system to the covalent modification site. The fact that there is no NOE from $A^{12}H2$ to AdoH3 implies that the CPI ring system is orientated 'edge-on' in the minor groove (Fig. 59). In agreement with this is the observation that $A^{12}H2$ shows medium intensity NOE connectivities with H8A2, H8A, H8B, H1B and H6 of adozelesins CPI headunit. Adozelesin CH_3 and H1A have slightly weaker intensity peaks, indicating that they are more distant from the adenine H2. The H8A2, H8A, H8B, H1B and H6 protons align the concave surface of adozelesin. The NOE data provide evidence of a close spatial relationship between these protons and $A^{12}H2$ situated at the base of the groove. This supports the proposed 'edge-on' binding conformation of the CPI headunit, where the convex surface of adozelesin is in close isohelical association with the base of the minor groove. If the remaining B and C subunits follow this trend (see later) then adozelesin binds analogously to (+)CC-1065, whose stereochemistry was inferred from the chirality of DNA (Reynolds *et al.*, 1985) as C8B(R) and C8a(S).

Other deoxyribose protons of A^{12}, C^{13} and T^3 also showed NOE connectivities with the A subunit. The H1′ protons are situated deeply within the minor groove so interaction with adozelesin is limited to the CH_3, which is also embedded deeply in the groove.

3.8 Noncovalent binding of adozelesin

Inspection of the NOESY spectra (both 400 and 600 MHz data sets) provided many drug-DNA interproton connectivities for defining the location of the indole (B) and benzofuran (C) subunits. NOEs between nucleotides A^4, A^5 and T^{11} to the indole (B) subunit, demonstrated that B lays across two base-pairs. The benzofuran subunit experiences an abundance of NOEs from T^{10} and C^9, with a few to G^6 and C^7, suggesting it also traverses two base-pairs. In all, adozelesin spans a total of five base-pairs. This is interpreted firstly from the NOEs occuring furthest in the 3′-direction along the covalently modified strand *i.e.* across $G^2 \equiv C^{13}$, where the tip of the CPI headunit is situated. Secondly, NOE connectivities to H5′2 and H6′2 can be used to pinpoint the location of the benzofuran subunit, which was found to be in between base-pairs $G^6 \equiv C^9$ and $C^7 \equiv G^8$.

3.8.1 The indole subunit

Both A^4H2 and A^5H2 display many NOEs to the H3′1 of the indole and the methylene H1A, H1B and methine H8A2 protons of the pyrrole. This implies that A^4H2 is situated between subunits A and B so that the indole covers part of A^4 and the whole of A^5 (Fig.59). These base H2 protons at the centre of the floor of the minor groove confirm, *via* their

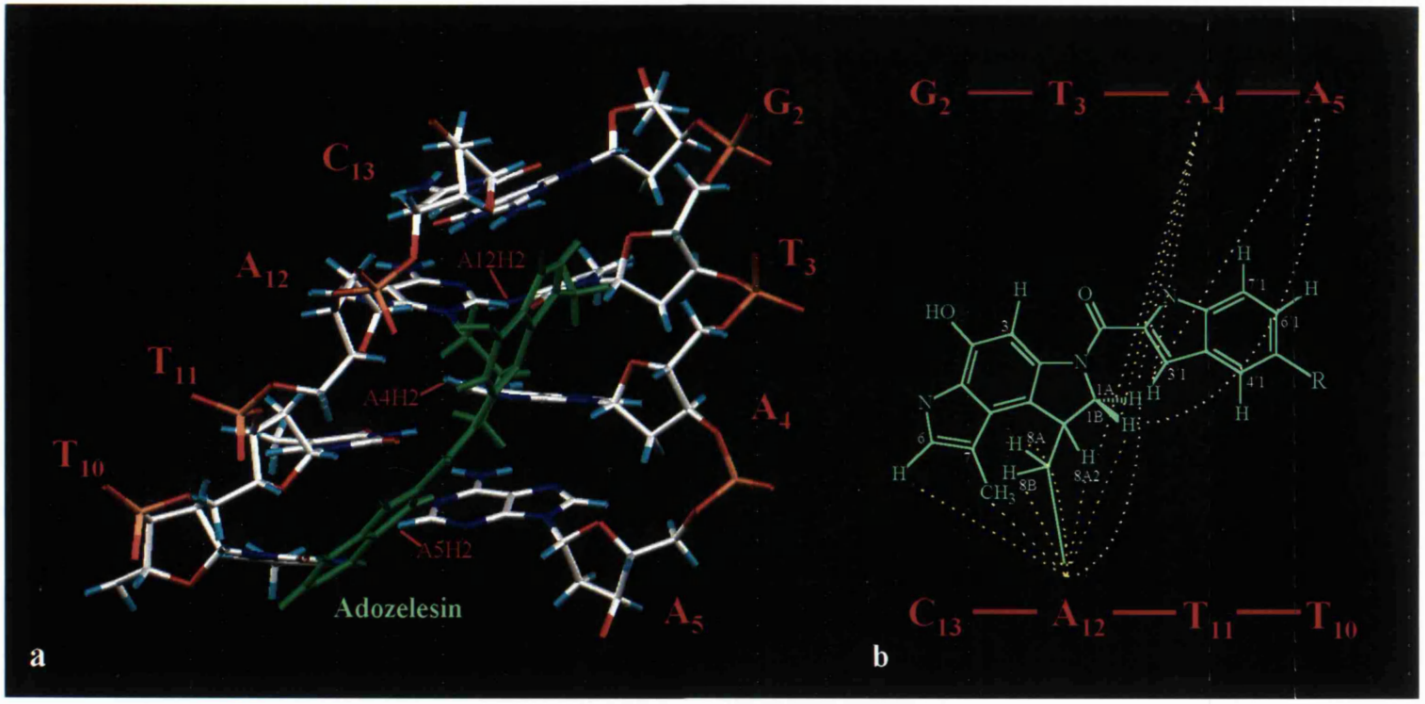


Figure 59 NOE connectivities between select adenine H2 protons and the bound adozelesin geminal H8A2, H1A, H1B, H8A and H8B protons. (A) Relative positions of the adenine H2 protons. (B) Relative intensities of the adenine H2 connectivities. (Intense peaks are shown in dashed yellow, intermediate peaks in dashed white).

connectivities to adozelesin protons H1A, H1B, H8A2 and H3′1, that the hydrophobic protons on the concave surface of the drug are situated deeply within the minor groove. Deoxyribose H1′ and H2′1 protons, which are also positioned low down on the wall of the groove, also display NOEs to the H3′1 and H4′1 of the indole. Meanwhile, the H4′, H5′1 and H5′2 sugar protons of T^{11} are responsible for NOE cross-peaks to H6′1 and H7′1, on the convex spine of the indole subunit of adozelesin. These sugar protons are located high on the walls of the minor groove, along the sugar-phosphate backbone. NOE connectivities from these protons to H6′1 and H7′1 but not to H3′1 or H4′1, suggest that the indole like the CPI subunit is orientated 'edge-on' within the minor groove. The hydrophobic, concave surface is isohelical with the floor of the minor groove whilst the hydrophilic, convex side of adozelesin points out of the groove and is exposed to the aqueous solvent. All the deoxyribose protons of T^{11} are shielded to some degree by the indole ring system. T^{11}H4′ is situated over the centre of the aromatic ring so exhibits a large upfield shift. The graph in Figure 52 clearly indicates the positioning of all three adozelesin subunits by the size of the deviation of chemical shift (-ve = shielding, +ve = deshielding). G^{2} and C^{13} experience some shielding by the CPI headunit, whilst A^{5} and T^{11} minima correspond to the indole and, lastly, the large dip at T^{10} correlates with the close proximity of the benzofuran. There are slightly more NOEs from the deoxyribose protons of the modified strand (8) to the indole subunit than the unmodified strand (5). This suggests that the indole is positioned parallel to both strands but experiences a marginally closer association with the modified, rather than the unmodified DNA strand.

3.8.2 The benzofuran subunit

There is a grand total of 26 NOE connectivities between the terminal benzofuran subunit and T^{10}, C^{9} and G^{8} of the covalently modified strand. Meanwhile, there are only 4 weak drug-DNA interactions with G^{6} and C^{7}, producing overwhelming evidence that the benzofuran is closely associated with the modified strand. C≡G base pairs have no aromatic H2 protons to help define the orientation of the ring system. However, there are a numerous contacts from the H1′, H2′1 and H2′2 protons in the base of the minor groove to H3′2, H4′2 and H5′2 but not H6′2 or H7′2, which verify the orientation of the benzofuran as 'edge-on'. The H4′, H5′1 and H5′2 protons of the sugar-phosphate backbone are high on the walls of the minor groove and see into H4′2, H5′2 H6′2 and H7′2 of adozelesin. This suggests, although the benzofuran ring is 'edge-on', it is slightly tilted so that all the protons on the aromatic ring system are close enough to result in an NOE. Measurement of the twist angle between the planes of the benzofuran and indole rings and the indole and CPI rings, yielded twist angles of -7·2° and +6·7° respectively (Fig. 61). The twist of these subunits relative to one another allows

adozelesin to yield a perfect complementary fit with the minor groove of β-helical DNA. The CPI and benzofuran subunits are nearly in plane, with only -0·5° between them.

Analysis of Figure 54 clearly shows $T^{10}H4'$ has the greatest upfield shift of 2ppm, closely followed by $T^{10}H5'2$ with 1·19ppm and $T^{10}H5'1$ with 0·73ppm. Medium intensity NOE cross-peaks occur between these protons and H4′2, H5′2, H6′2 and H7′2 protons of the benzofuran. These large upfield shifts are due to the strong shielding experienced from the six membered ring of the benzofuran, which these sugar protons are situated directly over (~2·5Å). In an analogous manner $C^{13}H5'1/H5'2$ are shielded slightly by the CPI headunit. Other protons of T^{10} such as H3′ and H2′1 are shielded only slightly because they are at the NOE distance limit for interaction at 5Å.

3.8.3 Sequence Selectivity

The indole (B) and benzofuran (C) subunits of adozelesin are responsible for the noncovalent binding mode of this drug. Noncovalent binding has been proved in many cases to be the source of sequence selectivity for binding of a drug to DNA, irrespective of whether covalent modification is involved. Adozelesin bound to 5′-d($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12*}C^{13}G^{13}$) at the preferred consensus sequence described by Weiland and Dooley, (1991) which is 5′-(T/A)(T/A)T A*(C/G)G. This also incorporates the 5′-(A/T)(A/T)A* sequence described by Yoon and Lee (1998). Hurley and Needham-VanDevanter (1986) attributed the binding specificity of (+)CC-1065 to the hydrophobic interactions between its concave surface and the adenine H2 protons. The van der Waals contacts between (+)CC-1065 and the narrow floor of the minor groove are proposed to drive the steric manoeuvreings to maximize the affinity of the ligand. These van der Waals forces hold non-polar molecules together, they are weak and of very short range, acting only between surfaces of molecules in close contact. Therefore, the van der Waals attractions between the macromolecule DNA (large surface area) and adozelesin are likely to be numerous, producing a significant force of attraction between these molecules.

Boger and Ishizaki (1990) have demonstrated that the biological activity of (+)CC-1065 is not determined by the electrophilic reactivity of the CPI headunit. Adozelesin, therefore, must alkylate DNA sequence-selectively by some other method. Boger *et al.*, (1990) examined the effect of removing the hydroxy and methoxy substituents of (+)CC-1065 (A) and (B) subunits. Only a small effect was noted on the binding affinity, while the AT-rich binding selectivity remained the same. Adozelesin, which does not have hydroxy and methoxy groups, has proved to be very AT-sequence-specific in alkylating 5′-d($CGTAAGCGCTTA^*CG)_2$ with

only one major product. Boger *et al.*, (1990; 1991a) established, through the synthesis of analogues of (+)CC-1065 with varying numbers of subunits, that subunits B and C are mandatory for achieving sequence-selectivity in the AT minor groove. Noncovalent interactions are responsible for the preferential binding to a specific consensus sequence (Boger *et al.*, 1991b). These noncovalent interactions are van der Waals forces and hydrophobic interactions which occur between adozelesin and the minor groove. Interactions such as these will be particularly prevalent in a narrower groove consisting of AT base pairs, instead of GC base pairs where the groove is wider. The increased width of a GC-rich groove leads to the drug associating more with one strand than the other, instead of yielding a close complementary fit with the drug (Garbesi *et al.*, 1996). Boger *et al.*, (1991b) found that 5′-d(TTA) was the second most preferred sequence after 5′-d(AAA) for (+)CBI-CDPI$_2$ alkylation, while 5′-d(TAA) was even less favoured. The 5′-d(TTA) consensus sequence must produce an environment within the minor groove through its base conformation, which encourages a greater number of interactions and hence stronger noncovalent binding than the 5′-d(TAA) sequence. It is likely that this is why adozelesin alkylates 5′-d($T^{10}T^{11}A^{12*}$) rather than 5′-d($T^{3}A^{4}A^{5*}$).

3.8.4.5 Hydrogen-bonding

Hydrogen-bonding in the case of many minor groove binders has been found to contribute only to the stability of the drug binding within the minor groove of DNA. Sequence selectivity has proved to be unconnected to hydrogen-bonding, for example: in the binding of oligopiptide antibiotics (Zakrzewska *et al.*, 1983); minor groove binding by the bis-benzimidazoles (Parkinson *et al.*, 1994); the noncovalent binding of the bis(quaternary ammonium) heterocycles (Gao *et al.*, 1993); the cyclopropapyrroloindoles (Hurley and Needham –VanDevanter, 1986).

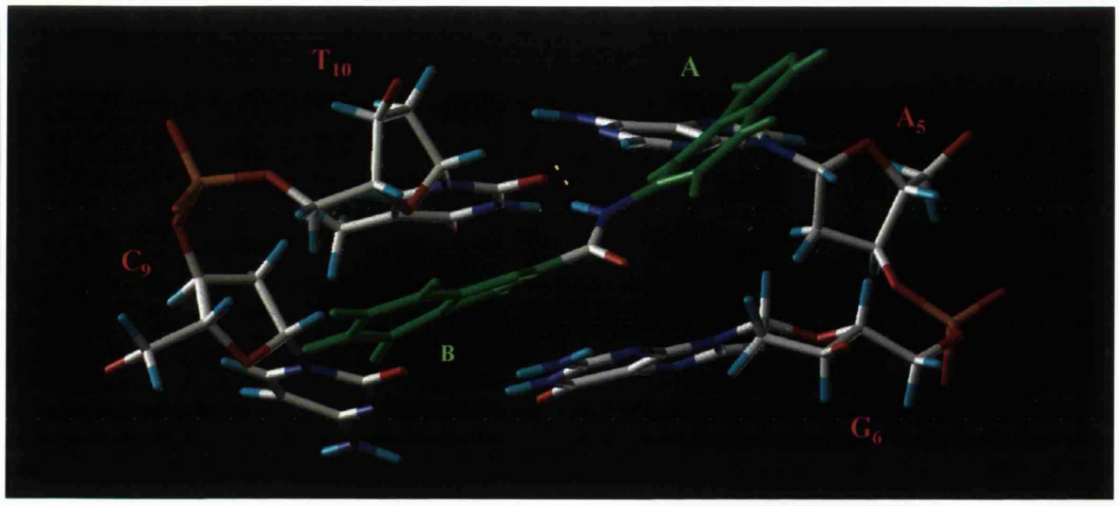


Figure 60 Illustration of the major hydrogen bond formed between the linking moiety and the carbonyl of thymine (T^{10}).

There are two direct hydrogen-bonds occurring within the adozelesin-5′-d(CGTAAGCGCTTA*CG) adduct. One hydrogen-bond is present from the NH of the linker between the indole and benzofuran subunits to the C2=O of T^{10} (Fig. 60). The short distance of 1·74Å indicates a strong pull of adozelesin towards the modified strand, which is verified by the large number of NOEs between 5′-d($C^9T^{10}T^{11}$) and adozelesin. The second hydrogen-bond arises between the CPI hydroxyl and C^{13}O1P, which is a phosphate oxygen. The cresent shape of adozelesin results in this –OH being very close (2·02Å) to the C^{13} 5′-phosphate group. Lin *et al.*, (1991) also noted a hydrogen-bond from the 8-phenolic proton of (+)CC-1065, however this was to the anionic oxygen of the phosphate two bases to the 3′-side of T, which constitutes the covalently modified base pair. The hydrogen-bonds within the adozelesin adduct are at either end of the drug-DNA complex, so stabilization takes place along the whole length of the adozelesin–DNA adduct.

3.8.5.6 Distortion around the G^6 base

Examination of the H2′1/H2′2→H6/H8 (as previously discussed in section 3.4.3.3) showed that there was a change in conformation around G^6, owing to the weakness of the NOEs connected with this nucleotide. G^6H8 is closer to its own 5′-flanking nucleotides A^5 sugar protons than its own. This could indicate more than a flip in sugar conformation from C2′endo (β-DNA) to C3′endo (A-DNA) is occurring. The presence of the base-sugar walk indicates that β-helical structure still prevails, though the weakness of the NOE connectivities shows that these interactions are at the distance limit. The refined molecular model generated from the NOE data indicates that there has been a rearrangement in the conformation of the sugar-phosphate backbone around G^6 (Fig. 61). This deformation would allow the deoxyribose sugar of G^6 to flip forwards so that this ring system was facing the sugar of A^5, giving rise to strong NOEs with A^5. The NOE data are consistent with a model like this. However, to determine whether the deformation is as described, torsion angles would be required from the NMR data and longer dynamics trajectories would need to be run (of the order of 1ns), to see if the macromolecular motion around this point would change on a longer timescale.

Seaman *et al.*, (1996) established that CPI-I bisalkylation of 5′-TAA^2AA^1 proceeded in a 'head to tail' fashion. This was attributed to a distortional wave focusing on a junction within the 5′-end TA step, on alkylation at A^1. This distortion is centred five base-pairs in the 5′-direction along from the modified base. DNA bending has also been described by Lin *et al.*, (see review by Hurley and Draves, 1993) on the covalent modification of 5′-d($G^1G^2C^3G^4G^5A^6G^7T^8T^9A^{10*}G^{11}G^{12}$)3′ by (+)CC-1065. Examination of the H8/H6→H1′ and H6/H6→H2′1/H2′2 intra and internucleotide connectivities demonstrated all nucleotides were

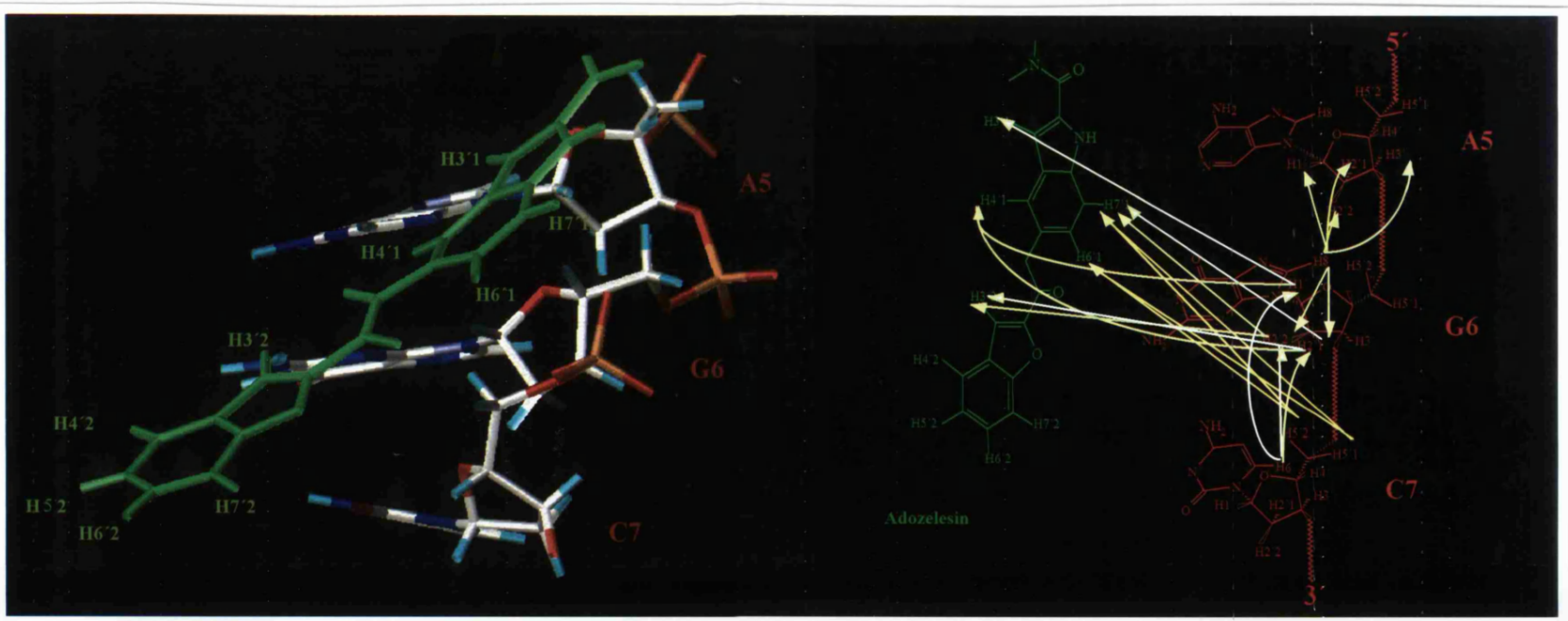


Figure 61 Section of the noncovalently modified strand of the adozelesin-DNA adduct which highlights the distortion around G^6.

in the anticonformation except for C^{18}. This cytosine is situated on the noncovalently modified strand, with its partner G^7 being 3 base-pairs upstream of the alkylation site at A^{10}. G^6 in the adozelesin-5′-d(CGTAAG^{6}CGCTTA*CG)$_2$3′ adduct is in an analogous position to this C^{18} and displays the same weakness in NOE connectivities. The distortion at C^{18} was attributed to the sugar assuming C3′-endo geometry, resulting from deformation of the backbone at A^{16} and A^{17} of the noncovalently modified strand on (+)CC-1065 binding. Like A^{16} and A^{17}, $A^5 \rightarrow G^6$ in this present study also displayed particularly intense NOEs on adozelesin binding. The observed distortion at G^6 could therefore be attributed to adozelesin binding locally at A^{12} on the opposite strand, or to a wave of distortion transmitted from covalent modification 6 base pairs downstream of G^6, on the same DNA strand.

Conformational changes around a C-G base pair have also been detected in homeodomain binding specificity studies (Ades and Sauer, 1994). Homeodomains are DNA binding motifs which have a preference for TAAT sequences. The wild type engrailed (segment polarity gene for *Drosophilia* development) homeodomain (from the *Drosophilia* transcription factor) is selective for an A:T flanking base pair at position 6 of the binding site, so the sequence TAATTA is more favoured over TAATCC. A CG base pair at positions 5 and/or 6 was found to cause conformational changes weakening interactions of the homeodomain residue. Crystal structures demonstrated that there were no contacts between the AT base pair and the homeodomain, so the sequence selectivity was induced on a DNA conformational basis. This evidence suggests that agents binding to the minor groove can induce distortive changes in DNA conformation between five and 6 base-pairs along the modified strand. It is possible that the observed distortion around this base (G^6) is the result of adozelesin alkylating A^{12}, followed by a distortional wave which centres around G^6 and C^7 of the modified strand.

Another hypothesis for the origin of this conformational change is the possibility of a weak, transient, bifurcating hydrogen-bond, from the linker between the B and C subunits of adozelesin to the guanine NH_2 encouraging distortion of this base. However, the refined molecular structure yields this hydrogen-bond distance as 4·3Å, owing to the migration of the benzofuran towards the modified strand. The NH of the linker produces a very solid H-bond to T^{10}, which will be preferred over the weaker one to G^6. No water molecule in this study was found to bridge a hydrogen-bond with G^6. It is likely that this hydrogen-bond pulls the drug ligand towards the modified strand, which stabilizes the benzofuran 'tail' so that it is not free to move around in the wider minor groove defined by the GC base-pairs. It is possible that the adozelesin may avoid steric crowding of the G-NH_2 by migrating towards the modified strand.

3.8 In summary

The cyclopropapyrroloindole adozelesin binds to the minor groove of the DNA sequence 5′-($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12}C^{13}G^{14}$)3′ (Cameron and Thompson, 1999). Alkylation took place on cyclopropyl ring opening, to the N3 of adenine (A^{12}) of the consensus sequence 5′-($T^{10}T^{11}A^{12}C^{13}$). This also indicated that both adozelesin molecules were bound in a 'tail to tail' orientation, yielding a self-complementary DNA adduct. The adenine base H2 proton was used to define the position of the CPI (A) headunit within the minor groove. Adozelesin was found to bind 'edge-on'. The convex, hydrophilic surface was directed out of the minor groove, while the concave, hydrophobic surface lay along the floor of this groove.

The benzofuran and indole subunits of adozelesin controlled the sequence selectivity element of binding. This sequence selectivity is a result of noncovalent interactions such as van der Waals forces and hydrophobic interactions between adozelesin and DNA. Hydrogen-bonds contribute to the stability of the complex. It is also clear that the interplay of the conformation of the flanking sequence and the topology of adozelesin is extremely important for adozelesin being able to squeeze tightly into the minor groove of DNA in a sequence-selective manner.

This study has proved that the cyclopropapyrroloindole adozelesin can achieve sequence selective binding to a DNA consensus sequence, which contains guanine and cytosine bases generally associated with wider DNA minor grooves. Adozelesin appears to only tolerate the presence of the CG base-pairs in the consensus sequence, because it can tether itself to the modified strand *via* a hydrogen-bond between the thymine carbonyl and the amide of the linker, so that the benzofuran is not able to 'flail' freely within the minor groove. This suggests an important role for the ureylene linker. Thompson *et al.*, (1995) have previously attributed the ability of bizelesin to interstrand cross-link the sequence 5′-TTAGTTA, to the hydrogen-bond formed between the ureylene linker C=O and the guanine-NH_2. The hydrogen-bond formed within this adozelesin-DNA adduct confirms the importance of the ureylene linker, in allowing this ligand to traverse an unpredicted consensus sequence which contains a guanine. Evidence of a distortion around the sixth flanking base-pair has also been provided, which may imply that CPI alkylation of adenine yields a unique distortional wave along the modified strand which culminates at G^6.

CHAPTER 4: RESULTS AND DISCUSSION

THE 5′d(CGAAAAACGG)·5′d(CCGTTTTTCG)-ADOZELESIN ADDUCT

The cyclopropapyrroloindole adozelesin like many other minor groove binders, has been shown to display a high sequence preference for poly 'A' tract sequences (Weiland and Dooley, 1991). Adenine-rich DNA is known to have special conformational character where the minor groove is deep and narrow, making it very attractive for molecules to bind within it, plus the helix axis is macroscopically bent (Nelson *et al.*, 1987). Although it has been established that the parent compound (+)CC-1065 binds covalently within the minor groove at the N3 of adenine *via* C8 (Hurley *et al.*, 1984), and Maxam-Gilbert sequencing has shown the preferred binding sequence to be 5′-AAAAA*, no (+)CC-1065-'A' tract DNA adduct has ever been published. Therefore, it was hoped that this study would yield an adozelesin 'A' tract DNA adduct, in order to elucidate the structure and why these monofunctional alkylating cyclopropapyrroloindole ligands prefer this binding sequence over all others.

The cyclopropapyrrdoindole adozelesin has also been studied on its ability to covalently modify 'A' tract DNA. The non-palindromic DNA sequence 5′d(CGAAAAACGG)·5′d(CCGTTTTTCG) was used in contrast to the shorter mixed A·T base sequence investigated in Chapter 3. Adozelesin bound to the DNA minor groove with a high degree of sequence selectivity to form a single drug-DNA adduct.

There are many steps to be undertaken sequentially before a refined molecular model of the 5′d(CGAAAAACGG)·5′d(CCGTTTTTCG)-adozelesin adduct (Fig. 62) can be achieved. These include: obtaining a pure DNA sample, accumulation and analysis of two-dimensional NMR data, production of a DNA-adozelesin adduct sample, collection and examination of a second 2D NMR data set, construction of a 3D computer molecular model of the adduct structure, followed by the application of NMR generated distance-range constraints to the structure *via* molecular mechanics and dynamics calculations. This process yields the refined molecular model of the adozelesin-DNA adduct.

4.1 Analysis of the 1D ^{1}H NMR spectrum

The 5′d(CGAAAAACGG)·5′d(CCGTTTTTCG) DNA duplex is non-self-complementary in nature. Therefore, its component strands have to be synthesized and purified seperately before they can be annealed together to form the required DNA duplex.

Figure 63 shows the one-dimensional NMR spectrum of the non-exchangeable protons within the 5′d($C^1G^2A^3A^4A^5A^6A^7C^8G^9G^{10}$) oligonucleotide strand. The spectrum is slightly broad, despite the inclusion of EDTA in the NMR buffer to chelate any paramagnetic ion contamination which can cause peak broadening. The spectrum was referenced to the D_2O peak at δ = 4·71ppm. No methyl singlets were noted between δ = 1 and δ = 2ppm, as was expected for a DNA strand deficient in thymine nucleotides. The adenine and guanine aromatic H8 protons, plus the cytosine and thymine aromatic H6 protons, are situated in the aromatic region between δ = 7·0 and δ = 8·25ppm. There appear to be nine peaks although the resonances at approximately δ = 7·4, δ = 7·7 and δ = 7·95ppm appear broad, suggesting overlaid proton signals. It is probable that, in an 'A' tract sequence consisting of a homologous adenine five base repeat, that the similarity of the chemical environment for each AH8 proton results in the superimposition of resonances. Five adenine H2 proton peaks are also overlaid with the H6/H8 aromatic protons in this region.

The 1D ^{1}H NMR spectrum of the sequence 5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) is given in Figure 64. D_2O is used as the external reference for this spectrum at δ = 4·71ppm. Inspection of the methyl region shows that there is a closely packed cluster of resonances correlating with the thymine methyl protons of $T^{14}\rightarrow T^{18}$, at approximately δ = 1·72, δ = 1·75, δ = 1·77, δ = 1·8 and δ = 1·82ppm. The aromatic region has a higher concentration of peaks at its upfield end towards δ = 7·5ppm in comparison with the spectrum in Figure 63 which corresponds to an homologous 'A' tract. This reflects the major pyrimidine content of the sequence where the base aromatic protons experience greater shielding, owing to weaker ring currents than in their purine counterparts (Scheek *et al.*, 1983). Singlets downfield at δ = 7·90 and δ = 7·95ppm are therefore most likely to be due to of G^{13} and G^{20}. The remaining set of overlapped peaks must therefore correspond to cytosine and thymine H6 protons.

After verification that the correct single strand nucleotide sequences were present in a pure form, the samples were mixed together and annealed to form β-helical duplex DNA. Hydroxylapatite chromatography was used to separate any excess single stranded DNA from the double stranded DNA.

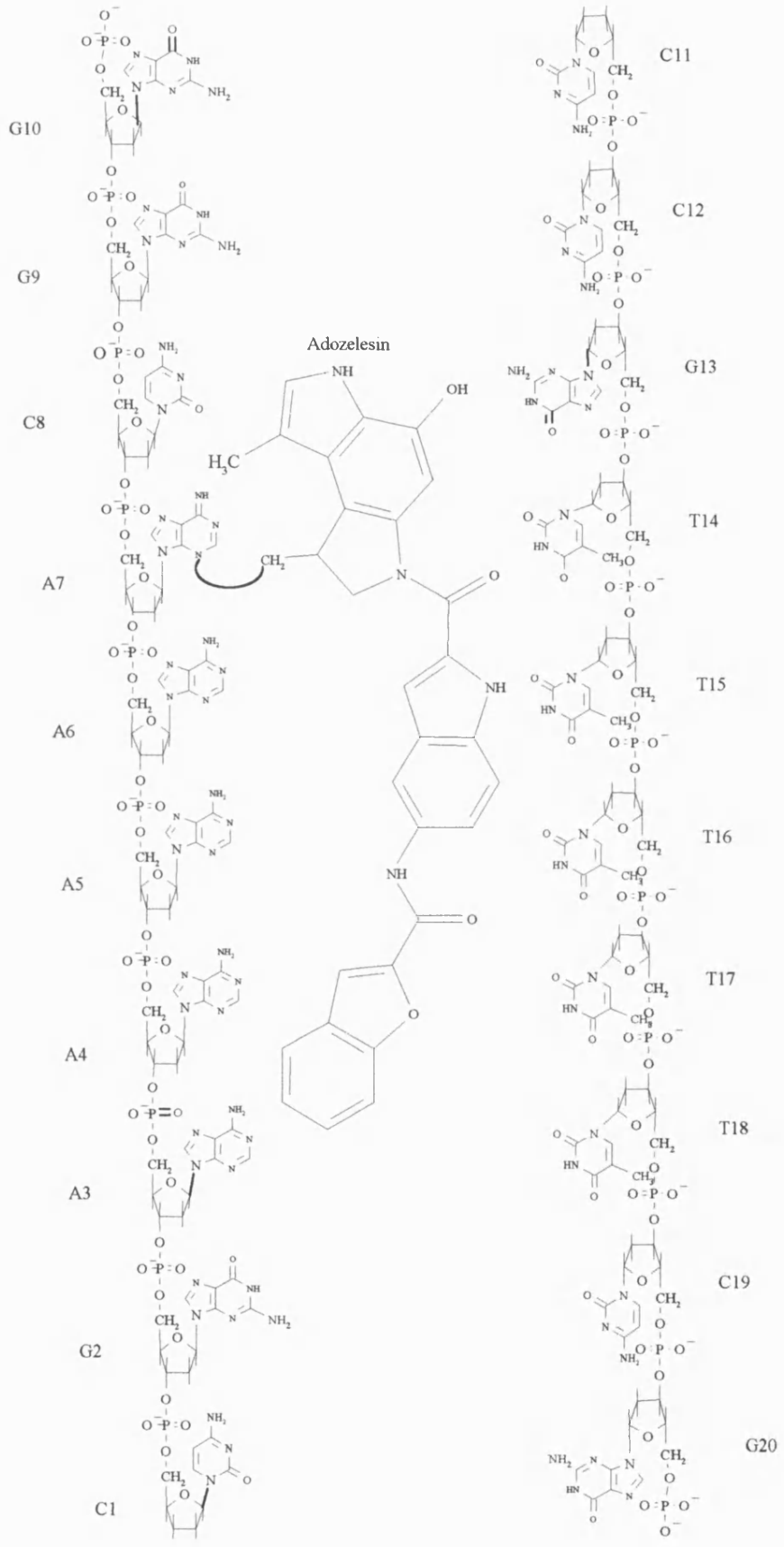


Figure 62. Schematic diagram of the adozelesin-DNA adduct.

The 1D ^{1}H NMR spectrum of the non-palindromic 5′d(CGAAAAACGG)·5′d(GCTTTTTGCC) duplex is given in Figure 65. The complexity of the spectrum has increased proportionally with the number of protons which now corresponds to two strands. On formation of the duplex, one thymine methyl has separated from the group shown in Figure 65, moving upfield from the cluster at δ = ~1·6 to δ = 1·4ppm. This is most likely to be a consequence of the duplex assuming straight 'A' tract character, with bends at either end on reverting to more general sequences of base pairing within the β-helical structure (Nelson *et al.*, 1987). 'A' Tract character is discussed in greater depth in section 4·4. Peaks in the aromatic region are less overlaid in the duplex than when the DNA is separated into its component halves. However, two broader peaks are still present at δ = 7·45 and δ = 7·85ppm, possibly corresponding to the 'T' and 'A' tracts respectively.

4.2 The 2D ^{1}H NOESY spectrum

Prior to reaction of the 5′d(CGAAAAACGG)·5′d(CCGTTTTTCG) DNA duplex with adozelesin, a 400MHz Jeol NOESY data set was accumulated (Appendix XIII Fig. 148). All non-exchangeable protons were then comprehensively assigned. It is very important to locate the chemical shifts of all protons in β-helical DNA before formation of the drug-DNA adduct. This allows comparison of their chemical shifts after the drug has reacted with DNA. Large chemical shift changes, either upfield or downfield, can indicate local changes in helical structure and, more importantly, give precise information on the exact positioning of the drug ligand within the minor groove.

This duplex DNA structure was more complicated to assign than was that of 5′d(CGTAAGCGCTTACG)$_2$, owing to the non-palindromic nature of the 5′d(CGAAAAACGG)·5′d(CCGTTTTTCG) sequence. Lack of self-complementarity meant that each oligomer constituting the DNA helix had a separate NOE connectivity walk. Great care had to be taken not to jump between proton connectivity walks within a region, which could easily lead to misassignment of the spectrum. This is quite difficult when each region is very crowded with peaks. The two-dimensional NOESY assignment strategies described in chapter 3 for analysis of the 5′d(CGTAAGCGCTTACG)$_2$ duplex were applied to this 'A' tract DNA. This utilized the assignment procedures documented by Wüthrich, (1986); Wijmenga *et al.*, (1993); Hare *et al.*, (1983); Chary *et al.*, (1987); Patel *et al.*, (1986); Weis *et al.*, (1984); Feigon *et al.*, (1983a,b); Scheek *et al.*, (1983;1984); Chazin *et al.*, (1986)

4.2.1 Cytosine H5 protons

The DNA duplex 5′d($C^1G^2A^3A^4A^5A^6A^7C^8G^9G^{10}$)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) contains five cytosine H5 protons. The cytosine H5 resonances are generally a good starting point for spectral assignment due to their high intensity, which makes them easy to identify. Five large cross-peaks were noted at δ 4·94, δ 5·18, δ 5·26, δ 5·57 and δ 5·65 ppm (Fig. 66). As previously discussed in section 3.2.1, the CH5 peaks can be distinguished by their correlations into their own base H6 proton and the base H6/H8 proton of the 5′-neighbour.

There are two terminal cytosines (C^1 and C^{11}) in this duplex which will have no correlations with a 5′-aromatic proton. The cross-peak at δ 5·57ppm is thus preliminarily assigned as C^1 or C^{11}. C^{11} is the 5′ neighbour of C^{12}H5. Examination of other peaks in the H6/H8→H1′ region, in the direction of H6 (δ 7·507ppm) from the C^1 or C^{11} cross-peak, yields only one other CH6→CH1′ cross-peak (δ 5·57ppm) verifying that it is a terminal. This peak has a cross-connectivity into a peak close beside it whose aromatic proton shift is δ 7·491ppm. This base proton also displays a cross-peak to CH5, which, in turn, shows an NOE to the -H6 of its 5′-neighbour. Two cytosine H5→H6 cross-peaks have therefore been assigned: C^{11}H5 δ 5·575ppm, C^{11}H6 δ 7·507ppm, C^{12}H5 δ 5·655ppm and C^{12}H6 δ 7·491ppm. The C^{19}H5, C^8H5 and C^1H5 remain to be assigned. Both of the highly intense cross-peaks at δ 4·937 and δ 5·258 have peaks positioned far downfield towards 8ppm with which they align. Only one cytosine (C^8H5) though has a 5′ neighbouring adenine (A^7) which would have such a low-field shift. C^{19}H5, in contrast, has a 5′ thymine (T^{18}) but no cross-peak between CH5 and T^{18}H6, which was expected to be located before δ 7·5ppm, was immediately apparent. The remaining cross-peak at δ 5·183ppm, as for the other peaks corresponding to CH5, seemed to align with peaks to 5′-neighbouring base protons as well as with a number of other peaks in the CH6 direction. No cross-peak corresponding to C^1H5 could therefore be definitively assigned, which corresponded to a terminal. To distinguish between C^1H5, C^8H5, and C^{19}H5 other regions of the spectrum *i.e.* H6/H8→H2′1/H2′2 and H6/H8→H1′ needed to be consulted, with specific examination of their cross-connectivity walks.

4.2.2 Thymine CH_3 protons

Observation of the methyl region in the 1D ^{1}H NMR spectrum (Fig. 67) revealed five methyl peaks corresponding to T^{14}→T^{18}, with chemical shifts around δ 1·4 to δ 1·6ppm. This eliminated the fairly intense cross-peak at δ 1·703ppm from assignment as a methyl, so, by

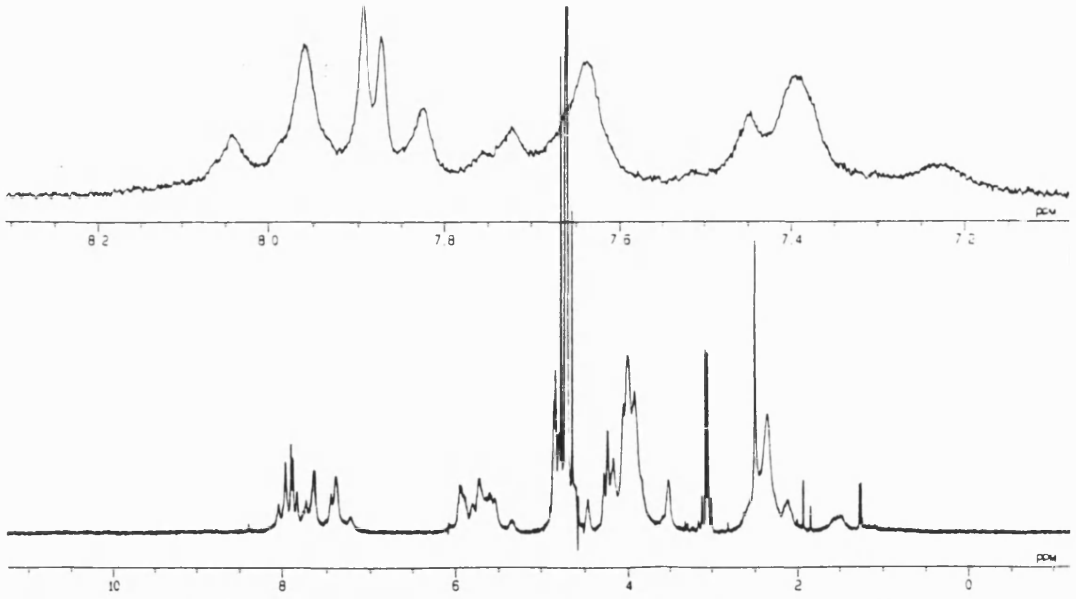


Figure 63. The 1D ^{1}H NMR spectrum of 5′d(CGAAAAACGG) single stranded DNA.

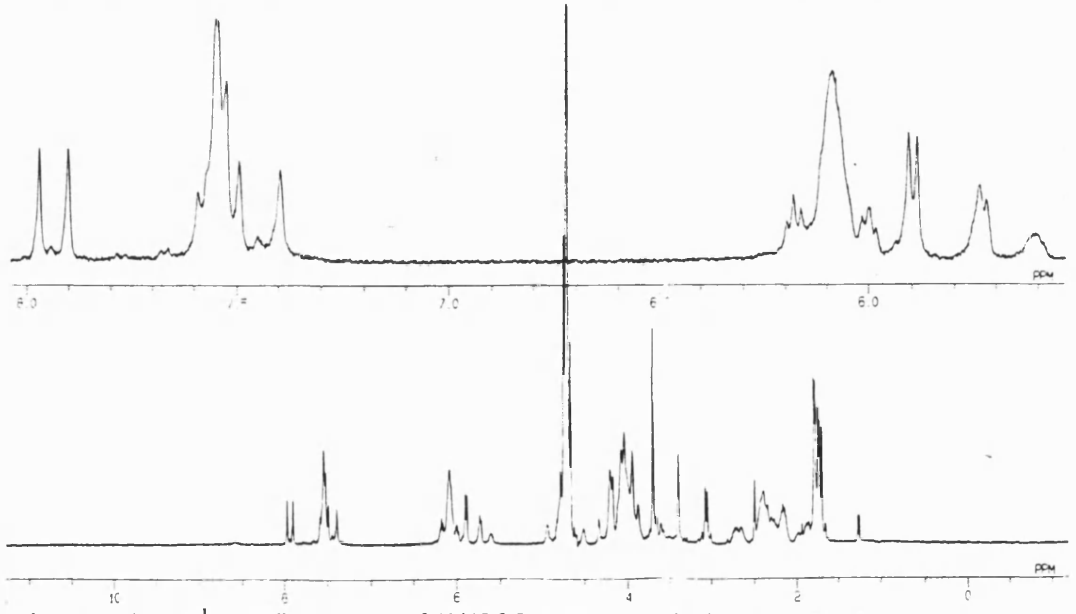


Figure 64The 1D ^{1}H NMR spectrum of 5′d(CCGTTTTTCG) single stranded DNA.

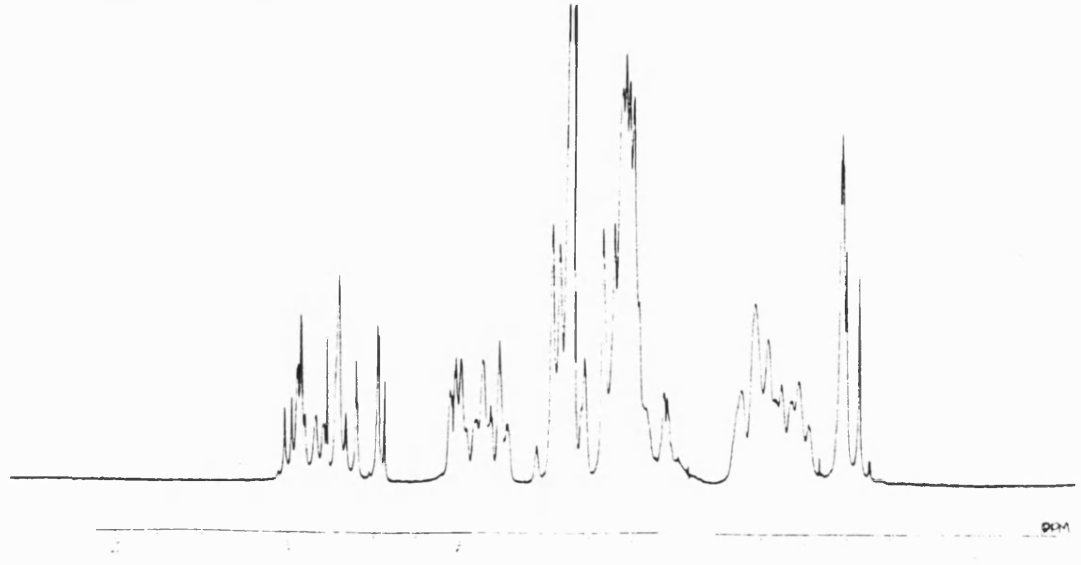


Figure 65 The 1D ^{1}H NMR spectrum of 5′d(CGAAAAACGG)·5′d(CCGTTTTTCG) DNA duplex.

default, it must belong to a sugar H2′1/H2′2 walk. The methyl peak at δ 1·367ppm was readily assigned as $T^{14}CH_3$, since it is the only resonance to have an intense correlation to a proton at δ 7·872ppm. Only a neighbouring 5′-guanine would produce such a low-field shift, so $T^{14}CH_3$ was assigned as δ 1·367ppm and $G^{13}H8$ as δ 7·872ppm. In the same manner, $T^{15}CH_3$ was assigned *via* its correlation to $T^{14}H6$. The remaining three methyls were in a tight cluster; so, to begin their assignment, a hypothesis was made on their positions using peak intensity as a guide. The resonance at the co-ordinates δ 7·402/1·54ppm appears to be much broader and is taller than the other methyls, possibly symbolizing the overlaying of peaks. If this peak was assigned as $T^{16}CH_3$ it would also be in the perfect position for a correlation with $T^{15}H6$, the aromatic proton of its 5′-neighbour. The methyl cross-peak at co-ordinates δ 7·349/1·575ppm seemed to consist of a main peak with a shoulder, which was also aligned with a large peak within the methyl cluster. This resulted in the preliminary assignment of $T^{18}CH_3$ at δ 1·593ppm, which correlated with $T^{17}H6$ at δ 7·349ppm. The cross-peak at the co-ordinates δ 7·349/1·575ppm was attributed to $T^{17}CH_3$ which showed a correlation with $T^{16}H6$ back into the main thymine methyl cluster. Of course, all of these hypothetical assignments (T^{16}, T^{17} and T^{18}) had to be duly validated by their NOE connectivities in other regions of the spectrum.

In contrast to the case of the 14-mer of DNA discussed in chapter 3, analysis of thymine methyl and cytosine H5 proton resonances did not produce conclusive evidence for exact assignment of chemical shifts. However, a number of good starting points for spectral assignment have been established: 5′d(_ _ _ _ _ _ _ _ _ _)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}$_ _) where _ denotes no proton assignment within the nucleotide. Little information has been gained on 'A' tract chemical shifts so far.

4.2.3 The aromatic→H2′1/H2′2 deoxyribose proton region

Assignment of the aromatic→H2′1/H2′2 region (Fig. 67) began by scrutinizing the spectrum for NOEs within a cross-connectivity walk, for which the aromatic base proton shifts had already been ascertained. The H2′1 and H2′2 protons were named using the SYBYL software notation for H2′ and H2′′ deoxyribose protons respectively. Beginning at $C^{11}H6$ (δ 7·507ppm), two NOE cross-peaks were located at δ 2·103 and δ 2·615ppm, corresponding to H2′1 and H2′2, although it was not possible at this stage to differentiate which peak tallied with H2′1 and which with H2′2. From these two NOE cross-peaks and knowing the position of $C^{12}H6$, the weaker internucleotide connectivities between $C^{11}H2'1/H2'2 \rightarrow C^{12}H6$ were located. The remaining peaks corresponding to $C^{12}H6$ were the stronger intranucleotide NOEs at δ 2·105 and δ 2·477ppm, relating to $C^{12}H6 \rightarrow C^{12}H2'1/H2'2$. All other protons were assigned in the

same manner moving sequentially along the strand. The methyl and aromatic proton chemical shifts for the thymines as proposed from the NOEs between their H6 and CH_3, proved to be correct. This also led to the assignment of $C^{19}H6$ to δ 6·963ppm and consequently to $C^{19}H5$ (δ 4·937ppm) *via* the NOE between them.

The 'A' tract assignment procedure began with examination of the H6 resonances which corresponded to C^1 and C^8H5. Within this assignment region there are only two peaks correlating to the H6 chemical shift of δ 5·183ppm, whilst four NOEs are apparent to the H6 at δ 5·258ppm. The former must be the terminal C^1 nucleotide, while the latter set are the strong intranucleotide NOEs from $C^8H6 \rightarrow H2'1/H2'2$ and the weaker internucleotide NOEs are $C^8H6 \rightarrow A^7H2'1/H2'2$. Two starting points are now available for beginning the 'walk', consisting of intra and internucleotide NOEs along the 'A' tract oligomer. Knowing the position of C^8 assisted 'A' tract assignment by exposing the direction of the adenine connectivities which are compacted close together.

4.2.4 The aromatic→H1′ deoxyribose proton region

This region was assigned in conjuction with the aromatic→H2′1/H2′2 region, to facilitate assignment when peaks became too overlaid or if internucleotide connectivities were excessively weak. The aromatic H6/H8→H1′ deoxyribose proton cross-connectivity walk in the 2D ^1H NMR NOESY spectrum is given in figure 66. The same principles of assignment were employed as described in section 3.2.4. After solving which CH5 peak belonged to which cytosine nucleotide with the aid of the aromatic H6/H8→H2′1/H2′2 spectrum, the chemical shifts of five CH6 resonances were also known. Analysis of the $TCH_3 \rightarrow H6/H8$ connectivities had produced three surely assigned aromatic chemical shifts for G^{13}, T^{14} and T^{15}, plus three more speculative ones for T^{16}, T^{17} and T^{18}. Assignment commenced at any of these protons, depending on where the peaks were most clearly resolved or if anything was known about the location of the previous or next nucleotide in the 'walk', in which case aligning their proton signals would give the 'walk' direction *e.g.* 5′→3′. The 'A' tract 'walk' proved to be easier to distinguish than that of the 'T' tract, where the clusters of peaks were very difficult to assign unambiguously. Here, consultation with the H1′→H2′1/H2′2, H6/H8→H3′ and H1′→H3′ regions, especially, to map the NOE cross-connectivities between them, was required to place all proton assignments on solid ground.

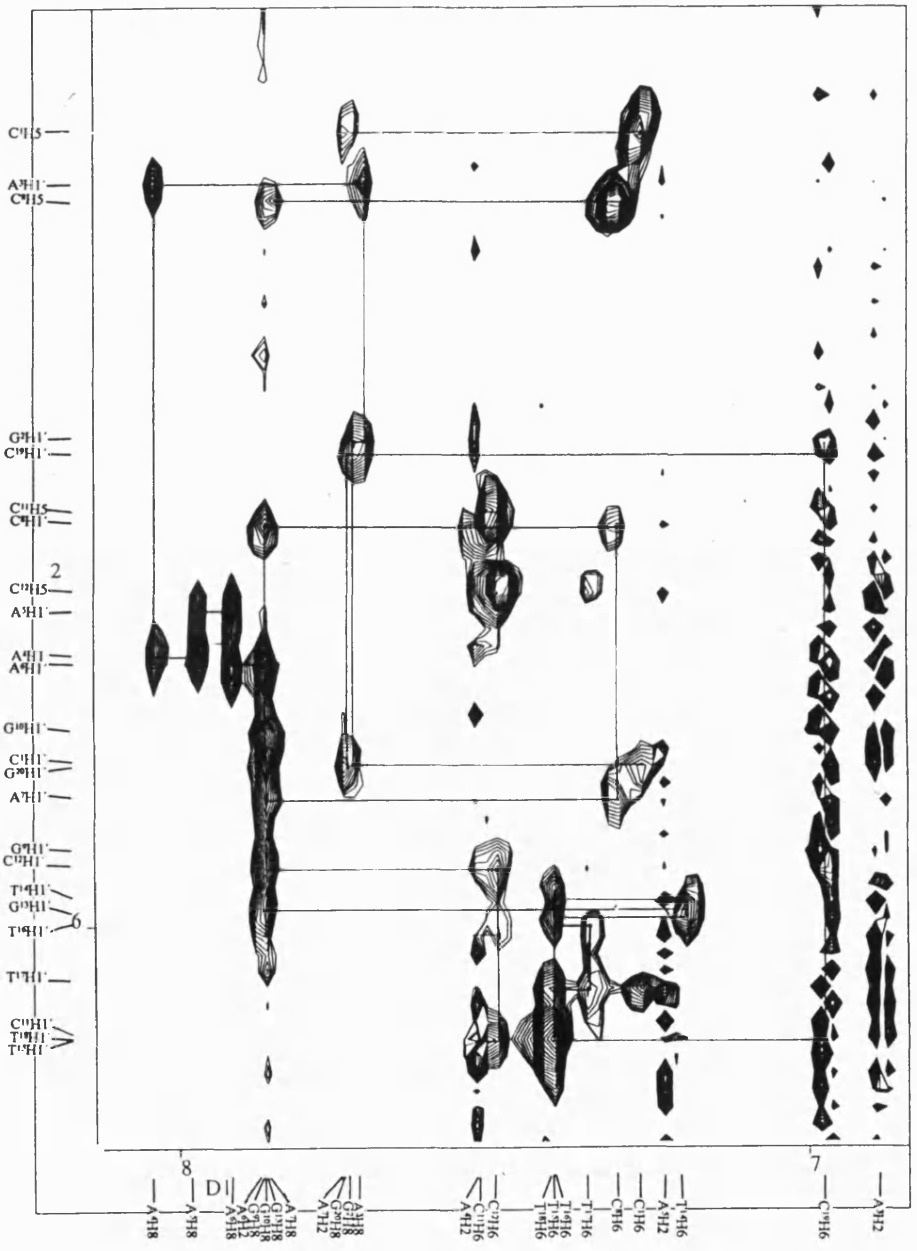


Figure 66 The H6/H8→H5/H1′ region of the 400MHz NOESY spectrum, of the 5′d($C^1G^2A^3A^4A^5A^6A^7C^8C^9G^{10}$)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA duplex

4.2.5 The H6/H8→H3′region

The H6/H8→H3′ 'walks' were located between δ 4·0 and δ 5·0ppm (Apppendix XIII Fig. 149). Chemical shifts of the 5′-terminal nucleotides of both the 'A' and 'T' tract strands were further upfield (C^1 δ 4·393 and C^{11} δ 4·461ppm) than those of the remaining nucleotide H3′ protons (δ 4·7→δ 5·0ppm). It is the presence of the 5′ terminal phosphate group with its negative charge, which is most likely to explain the upfield shift. Because the 5′-phosphate can rotate freely in solution, anisotropic effects around the carbonyl result in C^1 and C^{11} experiencing shielding of some of their protons and hence shifts to higher field (Chazin *et al.*, 1986). The H3′ 'walk' was nearly complete, with only a few internucleotide resonances missing. Identification of C^{19} resonances was quite difficult due to a t_1 ridge traversing the NOESY spectrum at this point. The H6/H8→H3′ proton 'walk' was assigned in the same manner as described in 3.2.5. All peaks were cross-referenced to the H6/H8→H1′ and H6/H8→H2′1/H2′2 regions to verify H6/H8 proton assignments, and the H1′→H3′ region to confirm H3′ assignments.

4.2.6 The H6/H8→H6/H8 region

The aromatic→aromatic region of the spectrum (Appendix XIII Fig. 150) maps internucleotide connectivities from base to base along the 5′d(CGAAAAACGG)·5′d(CCGTTTTTCG) oligomer strands. Referring to the proton chemical shifts derived from assignments of the H6/H8→H1′, H2′1, H2′2, H3′ regions, it was possible to jump from one base proton to another *e.g.* C^1H6(δ 7·263ppm)→G^2H8(δ 7·719ppm)→A3H8(δ 7·703ppm) *etc.* This area of the spectrum not only allowed verification assignments already made but greatly assisted on finding the route to a neighbouring nucleotide base in other complex and locally crowded regions of the NOESY spectrum.

Adenine H2 base proton NOE connectivites were also apparent in this region. Five AH2 trails were easy identifiable at δ 6·871, δ 7·172, δ 7·529, δ 7·755 and δ 7·861ppm, traversing the spectrum parallel to D2. A 'walk' between the AH2 protons was also possible from A^3 to A^5. It is quite difficult to decipher true peaks from the spurious noise of the AH2 track, so assignment was performed in conjunction with analysis of the H6/H8→H1′ region, using the procedure outlined in section 3.2.6. For example; A^5H2→A^5H1′(w), A^5H2→T^{16}H1′(w), A^5H2→T^{17}H1′(m) (where m = medium and w = weak intensity cross-peak). All duplex AH2 proton assignments are given in Table 5.

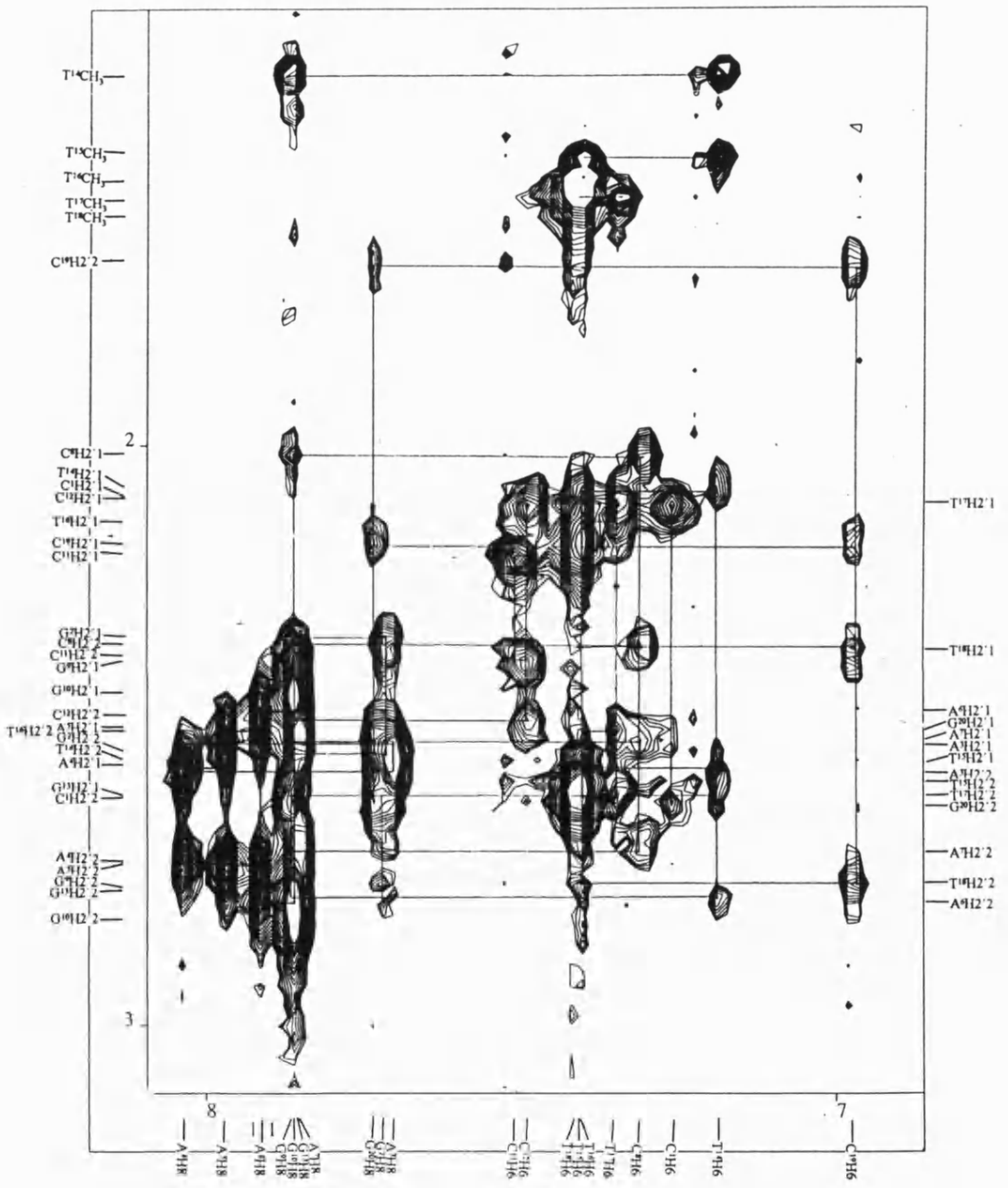


Figure 67 The aromatic→H2′1/H2′2/CH_3 region of the 400MHz NOESY spectrum of the 5′d(CGAAAAACGG)·5′d(CCGTTTTTCG) DNA duplex.

4.2.7 The H1′→H2′1/H2′2 region

Once all the H1′, H2′1 and H2′2 deoxyribose protons had been assigned using their NOE connectivities to aromatic protons, it was possible to check the correlation between them in the H1′→H2′1/H2′2 region (Appendix XIII Fig. 151). The H2′2 proton chemical shift was assigned to the most intense NOE of each pair, owing to the shorter interproton distance to H1′. Like the 5′(CGTAAGCGCTTACG)$_2$ duplex where the chemical shifts of most H2′1 protons were upfield of H2′2, this trend also predominated in this non-self-complementary duplex, however a higher proportion of H2′2 had chemical shifts upfield of those of the corresponding H2′1. This could indicate a slightly different conformation of β-helical DNA, due to the usual tendancy of H2′2 to be more deshielded by proton and pyrimidine ring currents than H2′1, hence leading to higher upfield shifting of H2′1 resonances.

4.2.8 The H6/H8→H4′ and H6/H8→H5′1/H5′2 region

Preliminary inspection of the H6/H8 aromatic proton to H4′, H5′1 and H5′2 deoxyribose proton NOE connectivities (Appendix XIII Fig. 149), indicated that the region δ 3·9→4·4ppm was very complicated, with many cross-peaks clustered close together in a small space making assignment difficult. Correlating cross-peaks in this region with those of H1′→H4′, H5′1, H5′2 was the only sure way to assign proton chemical shifts with certainty (Appendix XIII Fig. 152).

Only segments of the aromatic→H4′ 'walk', *e.g.* A3→A7, were present owing to the extensive spin-diffusion required for these internucleotide NOEs. H1′→H4′ yielded a more intense NOE cross-peak than H1′→H5′1/H5′2 resulting from the shorter interproton distance. This was a method to distinguish H4′ resonances from those of the H5′1/H5′2 deoxyribose protons. The H5′1 proton was separated from the H5′2 resonance according to Remin and Shugar's rule (1972), who stated that H5′1 resonates downfield of H5′2.

4.3 Summary of sequential resonance assignments

This concluded the assignment of the 400MHz NOESY proton NMR spectrum of the non-self-complementary duplex: 5′d(CGAAAAACGG)·5′d(CCGTTTTTCG). All chemical shifts of base and deoxyribose protons are displayed in Table 5. The presence of all sequential cross-connectivity walks between base aromatic and deoxyribose protons, both within the nucleotide (intranucleotide) and between nucleotides (internucleotide) within the same strand, proceeding in a 5′ to 3′ direction, verified this duplex was β-form DNA. The cross-peaks

relating to the methyls were also of similar intensity suggesting the base pairs were stacked evenly, resulting in nearly equal interproton distances, as would be expected for a right-handed β-helix. All pyrimidine H5 and CH_3 protons exhibited strong NOEs to the 5′ adjacent nucleotides H6 or H8, conforming to the requirements of β-DNA. It was not possible to determine whether the conformation of the deoxyribose sugar geometry was endo or exo, because this 400MHz data set was not of sufficient quality.

4.4 The unusual characteristics of 'A' tract DNA

The 5′d(CGAAAAACGG)·5′d(CCGTTTTTCG) DNA duplex contains a homologous 'A' tract repeat consisting of five adenine bases. There have been numerous studies on p(dA) DNA using a range of techniques such as NMR, gel electrophoresis and X-ray crystallography, in order to ascertain what special characteristics this sequence bestows on β-helical DNA.

Rhodes and Klug (1981) examined the helical periodicity of DNA in solution, using micrococcal nuclease and DNase I, which slice the oligonucleotide at the most accessible phosphodiester bond sites. Helical periodicity or the helical repeat is the number of base pairs required to complete a whole turn of the DNA helix. Analysis of the length of the resulting fragments showed that in DNA sequences with random permutations of nucleotides, 10·5 ± 0·1 base pairs (bp) are incorporated into each complete turn of DNA. In contrast poly(A) DNA sequences only had 10 ± 0·1bp constituting their helical repeat, suggesting a slightly different conformation in DNA structure.

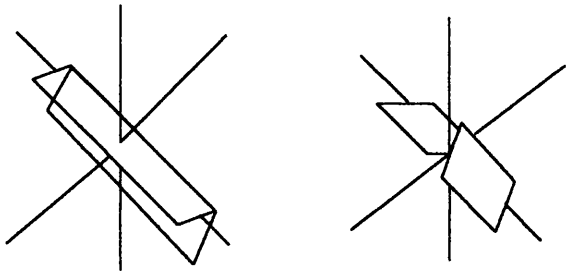


Figure 68 Schematic drawing of two base-pairs with roll and propellor twist (Wijmenga *et al.*, 1993)

An X-ray crystal structure of the DNA duplex sequence 5′d(CGCAAAAAAGCG) was studied by Nelson *et al.*, (1987). The 6bp long 'A' tract was found to be straight, with base pairs stacked parallel to each other and perpendicular to the axis of the helix. It is this straightness which typifies 'A' tract character. It was also established that these base pairs exhibit a high degree of propellor twisting about their longitudinal axis, the adenine often

rotating by about 20° with respect to its thymine partner (Fig. 68). This maximizes adenine-adenine stacking interactions between base pairs, presenting new hydrogen-bonding opportunities to stabilize this propellor twisted structure. Each propellor twisted adenine has its major groove facing edge, pointing downwards towards the 3′-terminal of its own strand. This results in the formation of an extra hydrogen-bond, in addition to Watson-Crick hydrogen-bonding, between the N6 amine of adenine and O4 of the thymine constituting the AT base pair to the 3′-side of adenine. Hydrogen-bonding between neighbouring base pairs comprises structurally stabilizing bifurcating bonds within the 'A' tract. In turn, this leads to a narrowing of the minor groove. Junction sites were also described at the 5′-pyrimidine-purine 3′-base step, where clashes between the propellor twisted bases and normal sequence DNA caused the base pairs to curl away from the minor groove. In contrast, the 5′-purine-pyridine junction site causes clashes between purines in the major groove, inducing base pairs to roll (Fig. 68) in the opposite direction. One of the most important findings described by Nelson *et al.*, (1987), was that helical twist between two halves of a helix turn are opposite in sign, so when they are adjacent they cancel each other out the helix axis remaining unchanged. However, when one of these halves was replaced with a normal sequence with no base pair roll, the 'A' tract roll of the other half-helical turn is not cancelled, resulting in the change in direction (bend) of the helical axis (Fig. 69).

Goodsell *et al.*, (1994) confirmed that the model Nelson proposed for macroscopic curvature of DNA was the correct one, resulting from the additive bending effects of alternate random and 'A' tract nucleotide sequences. They verified that tilt-wedge models, which attributed DNA bending to the rolling of base pairs caused by a junction site, were incorrect. TA steps were found to have a tendency to roll in a direction that compresses the minor groove. However, this was consistent with both wide and narrow minor grooves and not just 'A' tracts.

Katahira *et al.*, (1988) used NMR to examine the conformation of the 'A' tract sequence 5′d(CGCAAAAAAGCG). Two lines of evidence suggested that, although the helix was β-form, the 'A' tract bases were propellor twisted. Firstly, deuterium-exchange NOE data provided evidence of NOEs from thymine imino protons, not only to the adenine H2 to which it was base paired but also to the H2 of the 3′-adenine neighbour. Secondly, NOEs were noted between the adenine H2 and its own sugar H1′, as well as to the thymine H1′ of the 3′-neighbour on the complementary strand. In classical β-form DNA, the bases are not propellor twisted, so an NOE between this $A^{n}H2$ and $T^{n+1}H1'$ of the opposite strand would be extremely weak if observable at all. The intensity of the peaks was found to increase in the 5′→3′ direction along the strand, symbolizing the narrowing of the minor groove along the 'A' tract.

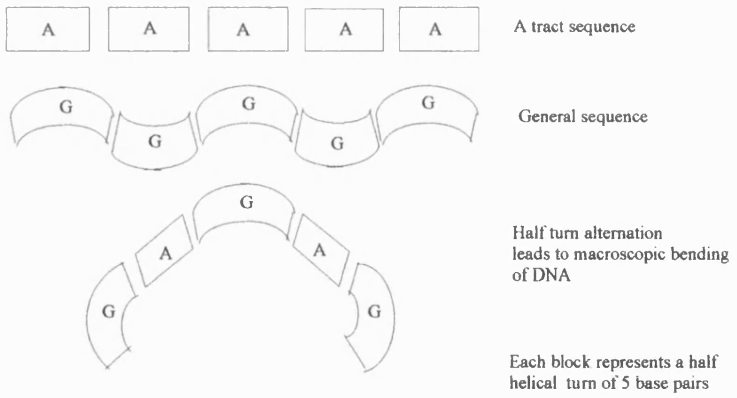


Figure 69 Schematic drawing demonstrating non-A-tract bending (Goodsell *et al.*, 1994; Nelson *et al.*, 1987)

Examination of the H6/H8→H1′ region of the 400MHz Jeol NOESY data for the sample 5′d($C^1G^2A^3A^4A^5A^6A^7C^8G^9G^{10}$)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$), also demonstrated that this DNA duplex exhibited 'A' tract character. For example, the NOEs between A^6H2→T^{16}H1′, A^5H2→T^{17}H1′ and A^4H2→T^{18}H1′ are stronger in intensity than intranucleotide NOEs to adenine H1′ or internucleotide NOEs to the thymine H1′ within the same base pair. This evidence supports the propellor twisting model of 'A' tract structure, with the adenine H2 being closer to the 3′-thymine neighbour of the thymine to which adenine is Watson-Crick base paired.

4.5 The Adozelesin-5′d(CGAAAAA*CGG)·5′d(CCGTTTTTCG) adduct

After the NMR analysis of the DNA duplex had been completed, the DNA was allowed to react with adozelesin. The resulting adduct was then subjected to further one- and two-dimensional NMR analysis. All 600MHz NMR data sets (NOESY, ROESY, TOCSY and COSY) were run courtesy of Dr. George Gray, at Varian, Palo Alto, C.A., U.S.A.

4.5.1 The 1D ^{1}H NMR spectrum

The one-dimensional proton spectrum in Figure 70 indicates that adozelesin has bound to the DNA duplex, to form the adozelesin-5′d(CGAAAAA*CGG)·5′d(CCGTTTTTCG) adduct. Superimposing this spectrum over the DNA duplex spectrum (Fig. 65) allowed the differences between the spectra to be clearly observed.

Observation of the methyl region showed that peaks due to the duplex had disappeared and, hence, the reaction between adozelesin and DNA had reached completion. Five methyl signals were immediately apparent at δ 1·347, δ 1·378, δ 1·466, δ 1·584 and δ 1·617ppm, which

would correspond to the thymine methyls at T^{14}, T^{15}, T^{16}, T^{17} and T^{18}. Adozelesin has only one methyl group on its CPI headunit at C7, with a shift slightly downfield at δ 2·07ppm (Walker *et al.*, 1997). There is no tall methyl singlet at this point, indicating that all free drug was removed during purification using a C_{18} Sep Pak. On preliminary analysis of the aromatic region, it seems not only very complex but possibly a little broad. However, this region (δ 6·8↔8·3ppm) contains 20 x H6/H8 and 5 x AH2 aromatic base protons attributable to the DNA, plus eleven adozelesin aromatic protons from its CPI, indole and benzofuran rings. This results in a grand total of 36 aromatic resonances in the small space of 1·5ppm and hence, the region would be expected to be crowded with peaks.

4.5.2 The 2D ^{1}H NMR DQF-COSY experiment

The initial step in the 2D ^{1}H NMR assignment strategy was to extract as much information as possible on scalar coupling from the 600MHz Varian DQF-COSY spectrum (Appendix XIV Fig. 153). Analysis of the phase-sensitive data for the adozelesin-5′d($C^1G^2A^3A^4A^5A^6A^7C^8G^9G^{10}$)·5′d($C^1C^2G^3T^4T^5T^6T^7T^8C^9G^{10}$) adduct, was completed in the same manner as previously described for the self-complementary adduct in section 3.4.2.

4.5.2.1 The CH5-CH6 region

Identification of the cytosine H6-H5 three bond vicinal couplings is always a good place to begin the assignment of non-exchangeable protons in a DNA containing sample. Appendix XIV Figure 154 displays the CH5(δ 5↔6ppm)-CH6(δ 6·8↔8ppm) region of the spectrum. The downfield end of the spectrum is perhaps a little out of phase in comparison to the upfield half, because there is a smearing effect over the large peaks so that they do not appear as an antiphase square array. Adjusting the phasing during processing using TRIAD did not irradicate this effect. However, even in this poorer quality spectrum, information about the locations of the CH5-CH6 cross-peaks was still extractable. A large discrete cross-peak was noted at CH5 δ 5·35, CH6 δ 7·03ppm. The signals of the remaining four cytosines are closely clustered together. The cytosine H5 cross-peaks at δ 5·51ppm(CH6 δ 7·385ppm) and δ 5·44ppm (CH6 δ 7·35ppm) are of equal intensity to the lone peak at δ 5·35ppm. This suggests that they are all equivalent to a single CH5-H6 correlation. Meanwhile, the CH5 peak at approximately δ 5·44ppm (CH6 δ 7·22ppm) is much greater in intensity than those previously described, indicating that two CH5 resonances are closely overlaid here.

The small cross-peak at δ 4·94ppm is attributable to the C^{19}H5 of unreacted DNA duplex. However, there are no similar COSY peaks attributable to C^1 and C^8H5-H6 resonances,

implying that the concentration of this species in solution must be very low, probably less than 10% of the population. There is also evidence of a second species in the sample. Three small CH5 cross-peaks were noted at approximately δ 5·45ppm (CH6 δ 6·9ppm), δ 5·75ppm (CH6 δ

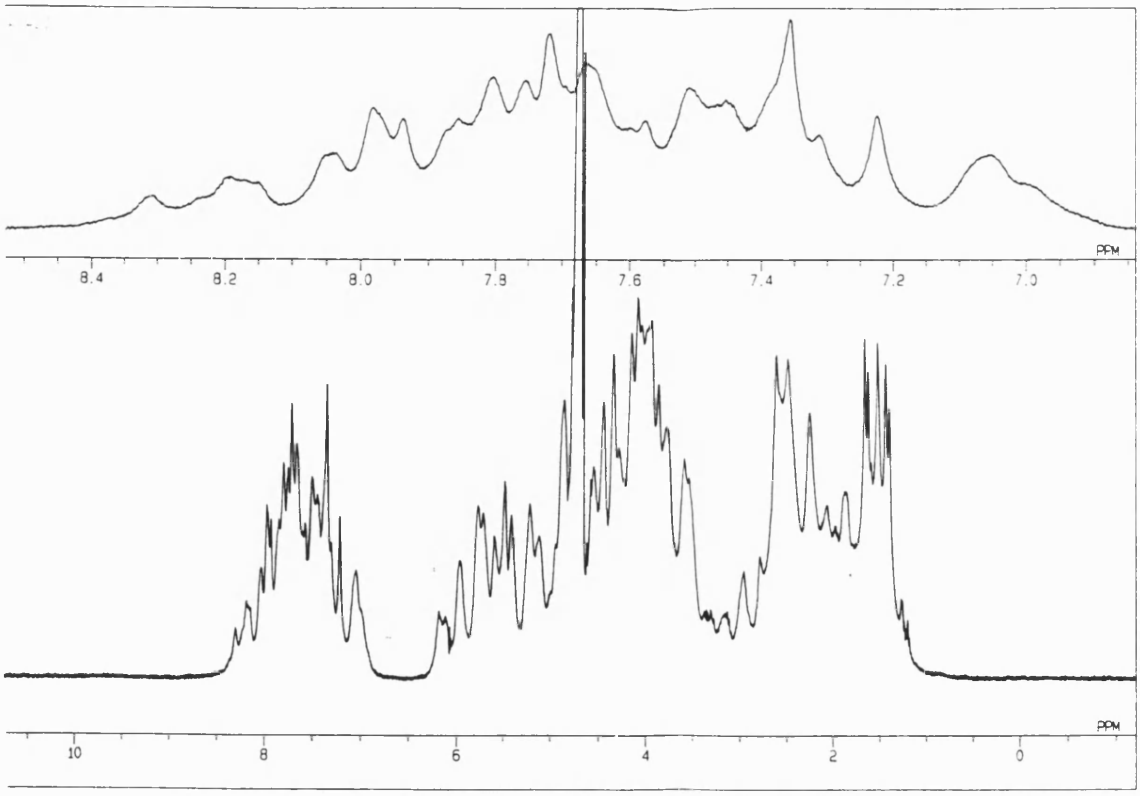


Figure 70 The 1D ^{1}H NMR spectrum of the adozelesin-5′d(CGAAAAA*CGG)·5′d(CCGTTTTTCG)adduct.

7·35ppm) and δ 5·90ppm (CH6 δ 7·55ppm). This second species could arise as a product of depurination or from alkylation of a second adenine site. Depurination occurs when the adozelesin-modified adenine (A^7) is removed from the DNA strand *via* β-elimination, which involves cleavage of the glycosidic bond, leaving the sugar without a base *i.e.* apurinic (Hurley *et al.*, 1984). Although this phenomenon can occur on standing over a long period of time, the adozelesin-DNA adduct of the self-complementary sequence was still in tact after 6 months suggesting it was stable. Therefore, it would follow that the 'A' tract DNA-drug adduct should also be stable. These second species cytosine peaks have also experienced quite a large chemical shift, more in keeping with an adozelesin bound to the minor groove of DNA, rather than a local effect bestowed on the DNA duplex by the loss of a cytosine base.

Alkylation of an adenine alternative to A^7 is a more favourable proposition, with A^6 being the most likely candidate for alkylation. A^6 would still have the preferred 5′ flanking sequence for adozelesin binding *i.e.* 5′d(A/T)(A/T)A* (Weiland and Dooley, 1991) which in this sequence is 5′d($A^3A^4A^5A^{6*}A^7$). Seaman *et al.*, (1996) have described the formation of a

monoadduct by CPI-I binding to 5′TAAAA*A rather than 5′TAAAAA*, where CPI-I is similar in structure to the first two subunits of adozelesin.

4.5.2.2 The TH6-TCH_3 region

Like the DQF-COSY spectrum of the self-complementary adozelesin-DNA adduct, no TH6-TCH_3 cross-peaks corresonding to a ^4J coupling were detected in this DQF-COSY spectrum. If the peaks were extremely weak, it is possible they were obscured by noise.

4.5.2.3 The deoxyribose sugar resonances

The areas of the DQF-COSY spectrum which corresponded to the spin-spin couplings around the deoxyribose ring were particularly crowded with COSY cross-peaks (Appendix XIV Fig. 155, 156, 157). Geminal two bond couplings (^2J) *e.g.* H2′1→H2′2 and H5′1→H5′2 yielded peaks close to the diagonal or possibly even obscured by it, when the chemical shift difference between these protons was approaching zero. Many ^3J couplings were also represented in the spectrum *e.g.* H1′→H2′1, H1′→H2′2, H2′1→H3′ and H3′→H4′. These cross-peaks were assigned in conjuction with the NOESY spectrum, verifying that the sequential NOESY assignments were correct.

4.5.2.4 The aromatic region

All scalar coupled protons giving rise to COSY cross-peaks within this region (Appendix XIV Fig. 154) correspond to the indole and benzofuran aromatic protons of adozelesin. These spin-spin couplings have already been described in Figure 49. The H6′1→H7′1 vicinal coupling was assigned to the cross-peak which showed no correlation to the other three, since it is an isolated spin system. Although it was assumed that H6′1 would be downfield of H7′1 (Walker *et al.*, 1997), the NOESY spectrum was required to confirm this by observation of the H6′1 weaker coupling to H4′1. The chemical shifts of the H6′1→H7′1 COSY cross-peak were at; δ 7·623 and δ 8·110ppm. The remaining three COSY peaks correlated to H4′2→H5′2, H5′2→H6′2 and H6′2→H7′2 of the benzofuran. Once again, the NOESY spectrum required consultation to give the direction of the connectivity network, by using the NOE H3′2→H4′2. These protons were found to have chemical shifts of: H3′2 δ 7·903, H4′2 δ 7·822, H5′2 δ 7·475, H6′2 δ 7·308 and H7′2 δ 8·000ppm.

4.5.3 The 2D ^{1}H NMR TOCSY spectrum

The 600MHz TOCSY data (Appendix XV) provided information on multi-step scalar couplings around the sugar ring *e.g.* H1′→H2′1/H2′2→H3′ *etc*, and around the benzofuran of adozelesin. This spectrum included the proton connectivities observed within the DQF-COSY spectrum, but this time they were clearly resolved unlike in the previous data set. As well as the ^2J and ^3J couplings seen in the DQF-COSY spectrum, this TOCSY spectrum displayed COSY connectivities on longer range scalar couplings such as ^4J and ^5J couplings.

Four large COSY peaks were noted at δ 5·43, δ 5·44, δ 5·52 and δ 5·35ppm. The intensity of the peak at approximately δ 5·43 suggested there were two proton signals clustered around the same point. All CH5 protons are then accounted for in the 5′d(CGAAAAACGG)·5′d(CCGTTTTTCG)-adozelesin adduct. The remaining low intensity peaks visible in this region are attributable to a small proportion of second species, which has already been discussed in section 4.5.2.1.

More peaks were visible in the aromatic region of the TOCSY spectrum than in the DQF-COSY. From the initial positions of the H4′2, H5′2, H6′2 and H7′2 protons of the benzofuran ring, it is possible to observe a succession of scalar connectivities:
H7′2→H6′2 is a ^3J coupling, strong COSY peak.
H7′2⟶H5′2 is a ^4J coupling, COSY peak is marginally weaker than the ^3J coupling.
H7′2⟶H4′2 is a ^5J coupling, COSY is less intense than the ^4J coupling.
These long range scalar couplings can be identified for all protons around the end of this benzofuran ring and also for H3′2→H4′2 and H3′1→H4′1. This is another method of confirming that these proton chemical shift assignments are correct. It also helps to simplify the NOESY spectral assignments because assignment of these cross-peaks, means they can be eliminated from becoming apart of the aromatic proton 'walks' of DNA which could lead to misassignments. Four peaks can be identified in the spectrum which do not correlate with any adozelesin proton assigned. These are most likely to be the most intense peaks of the second species which are present in the sample *i.e.* H6′1→H7′1, H4′2→H5′2, H5′2→H6′2 and H6′2→H7′2. This is because it is possible to walk between three of the peaks which are likely to correspond with the protons of the benzofuran ring. It is not possible to assign this second species because only a few extremely intense COSY peaks are visible relating to the drug, while an even smaller number will be present in the less sensitive NOESY experiment. Examination of the remaining regions of the spectrum provided confirmation of deoxyribose sugar proton assignments acquired from the two-dimensional NOESY spectrum.

4.5.4 The 2D ^{1}H NMR NOESY spectrum

Assignment of the 2D 600MHz ^{1}H NOESY spectrum (Appendix XVI Fig. 166) was accomplished using the approach, previously described for duplex DNA and the self-complementary adduct. Initial inspection of this spectrum reveals that it is exceedingly complicated with an abundance of DNA-DNA, adozelesin-adozelesin and DNA-adozelesin NOE connectivities in every region. Owing to the complexity of this spectrum, the DQF-COSY and TOCSY had to be consulted repeatedly to authenticate assignments. All interproton 'walks' within the adozelesin-5′d($C^1G^2A^3A^4A^5A^6A^7C^8G^9G^{10}$)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) adduct were complete except for the assignment of the protons of G^{10}, whose position was questionable, owing to its ability to correlate with a number of nucleotide resonances across the spectrum. However, its omission was not thought to cause a problem for generation of an accurate refined molecular model, because it is three base-pairs distant from the binding site. G^{10} is also a terminal nucleotide whose proton chemical shifts and peak intensities are not a true reflection of a guanine three base-pairs away from the binding site, because terminal base pairs have only one neighbour and often experience terminal end-fray effects.

4.5.4.1 The cytosine H5 region

The DQF-COSY and TOCSY spectra have already identified the chemical shifts of the five H5 cross-peaks at δ 5·35, δ 5·43(x2), δ 5·44 and δ 5·52ppm, which will correspond to the cytosines C^1, C^8, C^{11}, C^{12} and C^{19}. Examination of the most downfield CH5 shift at δ 5·52ppm in the NOESY spectrum (Fig. 71) showed there was a cross-peak directly underneath it which correlated with a peak to the right of it, and this peak was also in line with a cytosine cross-peak at δ 5·44ppm. This allowed a preliminary assignment of C^{11}H5 to δ 5·52ppm and C^{12}H5 to δ 5·44ppm, when the cross-peak at δ 7·385/5·612ppm corresponds to C^{11}H1′→C^{11}H6 and the one to its right is consistent with its 3′ neighbour C^{12}H6→C^{11}H1′. After the assignment of the C^{11} and C^{12}H5 cross-peaks, only those due to C^1, C^8 and C^{19} remained, two of which are nearly overlaid. Further analysis of these overlaid peaks shows that there are two lines of NOEs perpendicular to D2 which intersect these CH5 connectivities. C^8 is the 3′-neighbour of A^7; therefore, an NOE would be expected between C^8H6 and A^7H1′. If C^8H6 is taken to have a chemical shift of δ 7·186ppm, then there are a number of peaks at this chemical shift which have NOE connectivities to other protons δ< 7·5ppm but only one with an NOE correlation to a peak at δ 8·268ppm. A search for the most downfield NOE correlation to the CH5 at δ 5·436ppm only resulted in a cross-peak at δ 7·715/5·319ppm. A^7H8, being an adenine, is likely to be more deshielded than G^2H8; therefore, the CH5 at δ 5·436ppm was assigned as C^1H5, with C^1H6 and G^2H8 also now located. The cross-peak at co-ordinates δ 7·183/5·443ppm was assigned as δ 5·443ppm to C^8H5, with C^8H6 and A^7H8 being assigned as δ 7·183 and δ

8·268ppm, respectively. Hence, by default, the remaining CH5-CH6 NOE was attributed to C^{19} at δ 5·356ppm, which is the most upfield shifted cytosine as it was in the duplex DNA.

4.5.4.2 The thymine methyl region

The methyl connectivities of thymines T^{14}, T^{15}, T^{16}, T^{17} and T^{18} were assigned by seeking the same NOEs in Figure 72, as discussed in section 4.2.2. From C^{12}H6→H1′, it was possible to locate G^{13}H8, *via* interproton connectivity of G^{13}H8 to its 5′-flanking nucleotide C^{12}H1′. Knowing the location of G^{13}H8 was important for ascertaining the whereabouts of $T^{14}CH_3$ and the direction of the 'walk', between the thymine methyl and its 5′-neighbouring nucleotide H6/H8 base proton. A large NOE is visible in the methyl region at G^{13}H8 which shows only one correlation into a methyl; therefore, this must be $T^{14}CH_3$ δ 1·617ppm and T^{14}H6 δ 7·454ppm. Next, looking for the internucleotide correlation between $T^{15}CH_3$ and T^{14}H6 will provide the chemical shift of the next methyl in the sequence, *i.e.* δ 1·466ppm, plus the assignment of T^{15}H6 δ 7·417ppm. The remaining methyls were assigned in the same manner and their shifts are given in Table 5.

4.5.4.3 The H2′1/H2′2→H6/H8 region

The chemical shifts of the base protons already assigned are: 5′d(C^1G^2_ _ _ _ A^7C^8_ _)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}$_), where the _ denotes unassigned nucleotides. Beginning at any of these sites, it is possible to move forwards or backwards along the DNA strand, sequentially assigning the strong H6/H8 intranucleotide connectivities to H2′1 and H2′2, plus weaker internucleotide NOEs to H2′1 and H2′2 of the 5′ neighbour. All chemical shifts of H2′1 and H2′2 protons from this 'walk' (Fig. 72) were entered into Table 5.

4.5.4.4 The H1′, H2′1, H2′2, H3′, H4′, H5′1, H5′2 proton connectivities

Sequential assignment of the deoxyribose protons H1′, H3′, H4′, H5′1 and H5′2 were made previously described for the duplex DNA (Appendix XVI Fig. 165-171). Owing to the complex nature of this spectrum, all areas of the spectrum had to be consulted to confirm the chemical shift of each proton. With geminal and vicinally coupled protons, the NOE cross-peaks were checked to see if they corresponded to the COSY cross-peaks of the DQF-COSY and TOCSY spectra. H4′, H5′1 and H5′2 resonances were stronger and clearer on their correlation to H1′ and H3′, where interproton distances were shorter and the effects of spin diffusion were more pronounced than

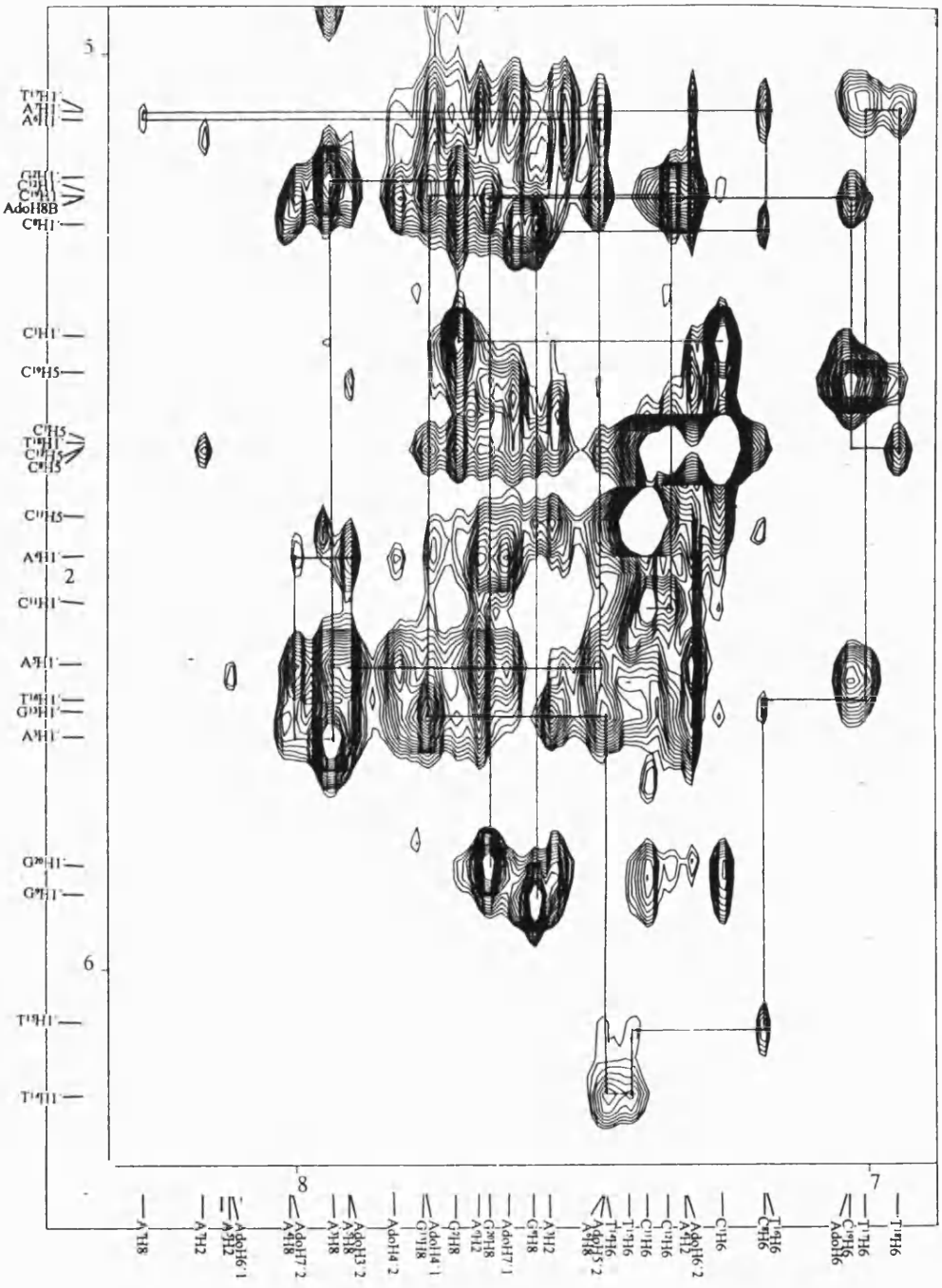


Figure 71 The H6/H8→H1′/H5 region of the 2D ^{1}H NOESY spectrum of the adozelesin-5′d($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7}C^{8}G^{9}G^{10}$)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct.

in their NOE connectivities to H6/H8. When all protons of each nucleotide had been assigned within their own region, resulting in complete proton connectivity 'walks' (except for G^{10}) and these local assignments then mapped correctly with the assignments in all other areas of the spectrum, the assignment procedure was complete. The cross-referencing of all proton chemical shifts placed their assignment on solid ground. The presence of all proton 'walks' confirmed this 'A' tract had maintained β-helical structure.

4.5.4.5 The adenine H2 protons

Owing to the complexity of the spectrum resulting from the non-palindromic nature of the duplex, plus slight line broadening which had a detrimental effect on the resolution, the AH2 protons were quite difficult to locate. However, after inspection of both this spectrum at a lower contour slice (closer to the baseline) and the tall, sharp resonance lines in the 1D spectrum, the adenine H2s were identified. Each produced a faint trail of NOEs which were particularly prevalent in the H6/H8→H1′ region. A connectivity 'walk' from $A^3 \rightarrow A^5H2$ was located in the aromatic region, verifying assignments. Like the adenine H2 protons of the duplex DNA, the strongest cross-strand NOEs were to the 3′-neighbouring thymine to which the adenine was base-paired *e.g.* $A^5H2 \rightarrow T^{17}H1'$ and $A^4H2 \rightarrow T^{18}H1'$ (Fig. 71). This indicated that the 'A' tract had retained its propellor twisted character (Katahira *et al.*, 1988), even after adozelesin had bound to the minor groove of DNA to form an adduct.

This concluded the assignment of the non-exchangeable sugar and base protons of DNA. All proton chemical shifts are given in Table 5. The chemical shift difference was calculated for each deoxynucleotide by subtracting the duplex value from the adduct value. The change in chemical shift was then represented graphically (Fig. 73) to illustrate which protons experienced the greatest shifts on the addition of adozelesin to DNA. These sequential connectivity networks, with their distinct patterns of peak intensities along each DNA strand, indicated β-helical structure with Watson-Crick base-pairing and straight 'A' tract character had been preserved on drug-DNA adduct formation.

4.5.5 ^{1}H NMR assignments of the adozelesin non-exchangeable protons

The aromatic ring protons of the indole and benzofuran subunits of adozelesin have already been assigned using the DQF-COSY and more comprehensive TOCSY NMR data. COSY peaks resulting from the *ortho* couplings between H4′2, H5′2, H6′2 and H7′2 of the

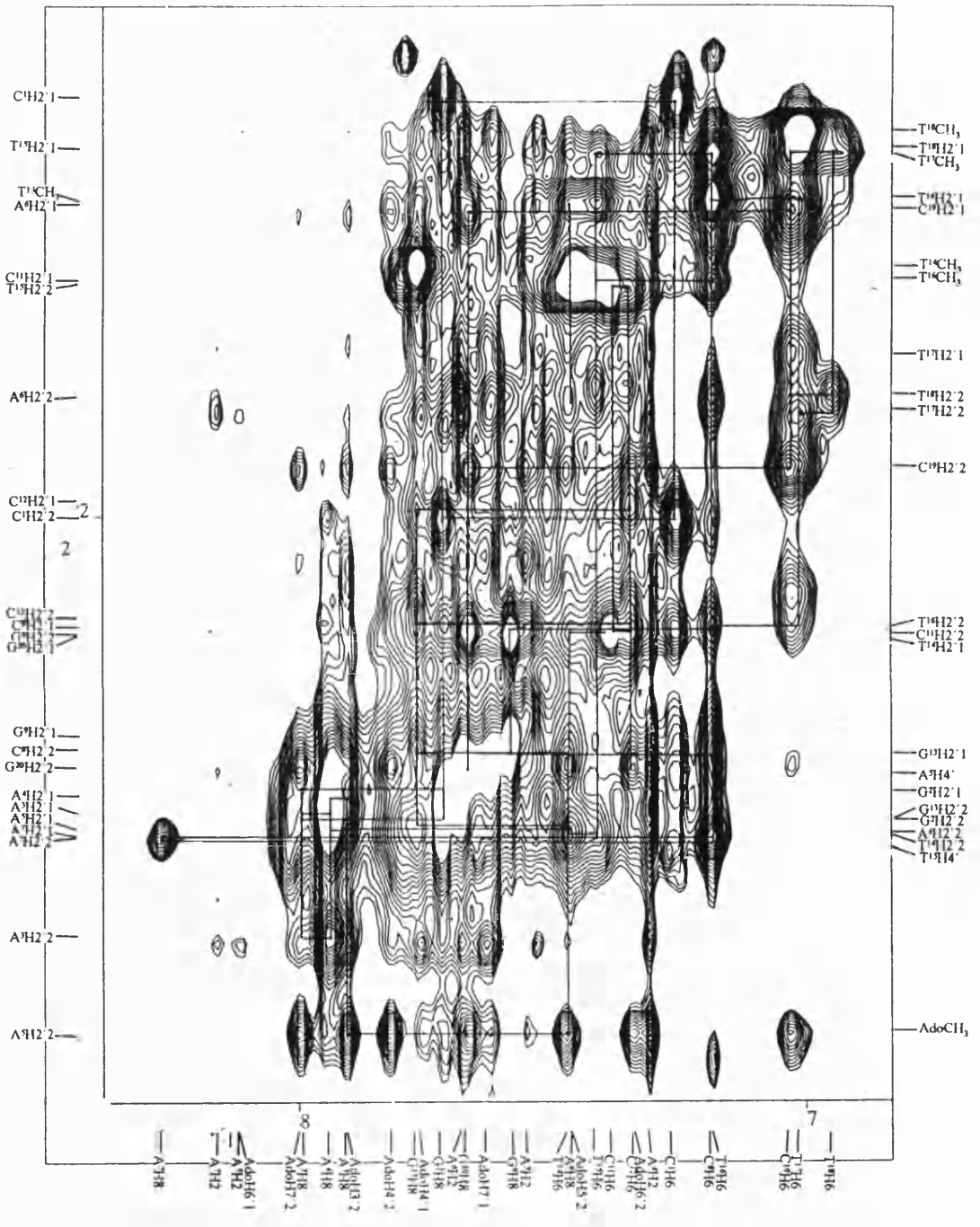


Figure 72 The H6/H8→CH_3, H2′1, H2′2 of the 2D ^{1}H NOESY spectrum of the adozelesin-5′d($C^1G^2A^3A^4A^5A^6A^7C^8G^9G^{10}$)·5′d′($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct.

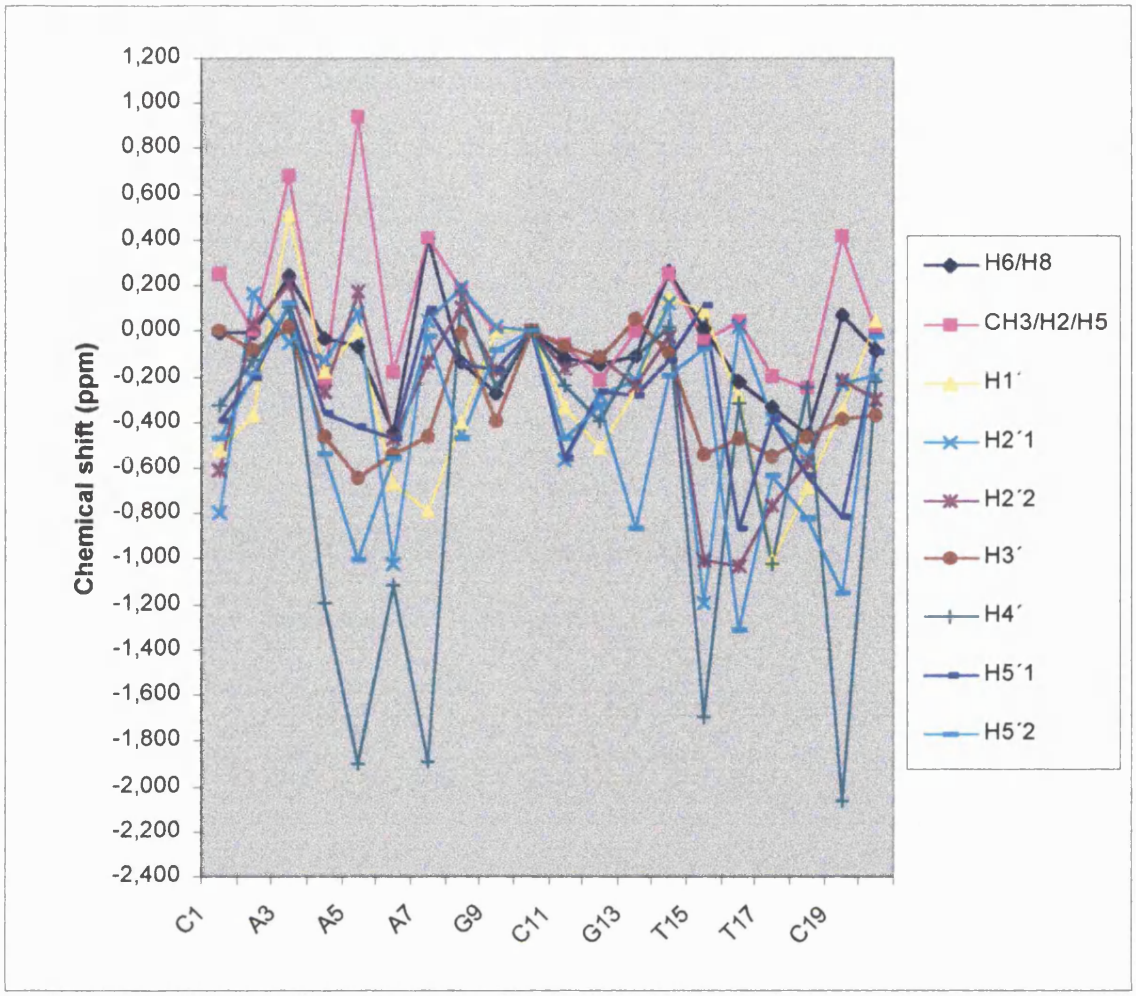


Figure 73 Graphical representation of the chemical shift changes that occur for each proton in DNA, as a result of the covalent binding of adozelesin to form an adduct.

benzofuran, plus H6′1 and H7′1 of the indole, were noted in both the DQF-COSY and the TOCSY spectra. Longer range ^{4}J couplings, for example H3′1→H4′1 and H3′2→H4′2 were only identified from the TOCSY data. All these proton assignments were confirmed by their through-space connectivity 'walks', present in the aromatic→aromatic region of the NOESY spectrum (Appendix XVI Fig. 166). The chemical shifts of the B and C subunits of adozelesin were entered into Table 5.

This left only the protons of the CPI headunit to be located in the NOESY spectrum. Using the same reasoning outlined in section 3.4.4, the H8A, H8B, H8A2, H1A, H1B, H3, CH_3 and H6 protons were assigned and their chemical shifts entered into Table 3B. Strong COSY correlations between H8A→H8A2 (δ 3·73/4·30ppm) and H1A→H8A2 (δ 3·71/4·30ppm) were clearly visible in the TOCSY spectrum (Appendix XV Fig. 164).

Chemical shifts for the non-exchangeable protons of DNA in the DNA duplex and Adozelesin-d(CGAAAAACGG).d(CCGTTTTTCG) adduct.

	H6/H8			CH3/H2/H5			H1'			H2'1			H2'2			H3'			H4'			H5'1			H5'2		
C1	7,263	7,257	-0,006	5,183	5,436	0,253	6,05	5,319	-0,75	2,103	1,308	-0,795	2,615	2,010	-0,605	4,393	4,392	-0,001	4,159	3,833	-0,326	3,955	3,561	-0,394	3,956	3,481	-0,475
G2	7,719	7,715	-0,004				5,504	5,139	-0,365	2,337	2,504	0,167	2,525	2,557	0,032	4,893	4,812	-0,081	4,257	4,128	-0,129	4,024	3,816	-0,208	3,928	3,744	-0,184
A3	7,703	7,945	0,242	6,871	7,551	0,680	5,235	5,744	0,509	2,571	2,520	-0,051	2,517	2,727	0,210	4,885	4,905	0,020	4,201	4,306	0,105	3,999	4,119	0,120	3,933	4,055	0,122
A4	8,036	8,001	-0,035	7,529	7,307	-0,222	5,730	5,551	-0,179	2,563	2,435	-0,128	2,718	2,451	-0,267	4,961	4,498	-0,463	4,305	3,111	-1,194	4,089	3,729	-0,360	4,001	3,457	-0,544
A5	7,968	7,904	-0,064	7,172	8,111	0,939	5,671	5,672	0,001	2,457	2,538	0,081	2,723	2,896	0,173	4,945	4,299	-0,646	4,325	2,424	-1,901	4,139	3,718	-0,421	4,112	3,109	-1,003
A6	7,915	7,473	-0,442	7,861	7,679	-0,182	5,741	5,072	-0,669	2,492	1,473	-1,019	2,289	1,816	-0,473	4,943	4,401	-0,542	4,341	3,226	-1,115	4,192	3,723	-0,469	4,132	3,577	-0,555
A7	7,863	8,268	0,405	7,755	8,159	0,404	5,869	5,078	-0,791	2,515	2,562	0,047	2,703	2,565	-0,138	4,762	4,304	-0,458	4,343	2,450	-1,893	4,092	4,188	0,096	4,046	4,026	-0,020
C8	7,315	7,183	-0,132	5,258	5,443	0,185	5,595	5,194	-0,401	2,015	2,209	0,194	2,357	2,461	0,104	4,801	4,798	-0,003	4,011	4,187	0,176	4,109	3,962	-0,147	4,019	3,550	-0,469
G9	7,860	7,583	-0,277				5,926	5,925	-0,001	2,367	2,387	0,020	2,378	2,212	-0,166	4,908	4,514	-0,394	4,358	4,110	-0,248	4,212	4,045	-0,167	4,036	3,957	-0,079
G10	7,861						5,823			2,836			2,437			4,935			4,371			4,180			4,144		
C11	7,507	7,385	-0,122	5,575	5,515	-0,060	6,141	5,612	-0,329	2,188	1,622	-0,566	2,372	2,209	-0,163	4,461	4,399	-0,062	4,147	3,908	-0,239	4,081	3,521	-0,560	3,937	3,463	-0,474
C12	7,491	7,346	-0,145	5,655	5,442	-0,213	5,942	5,432	-0,510	2,477	2,198	-0,279	2,105	1,996	-0,109	4,767	4,652	-0,115	4,297	3,908	-0,389	4,107	3,840	-0,267	4,085	3,763	-0,322
G13	7,872	7,766	-0,106				5,982	5,724	-0,258	2,615	2,406	-0,209	2,788	2,549	-0,239	4,797	4,851	0,054	4,335	4,222	-0,113	4,220	3,941	-0,279	4,112	3,248	-0,864
T14	7,191	7,454	0,263	1,367	1,617	0,250	6,069	6,142	0,150	2,085	2,206	0,121	2,565	2,546	-0,019	4,921	4,827	-0,094	4,220	4,241	0,021	4,212	4,085	-0,127	4,205	4,010	-0,195
T15	7,403	7,417	0,014	1,505	1,466	-0,039	6,122	6,061	0,092	2,580	1,383	-1,197	2,620	1,614	-1,006	4,847	4,304	-0,543	4,257	2,566	-1,691	4,137	4,249	0,112	4,087	4,016	-0,071
T16	7,402	7,179	-0,223	1,540	1,584	0,044	6,002	5,708	-0,294	2,171	2,190	0,019	2,492	1,461	-1,031	4,857	4,389	-0,468	4,290	3,973	-0,317	4,308	3,444	-0,864	4,223	2,908	-1,315
T17	7,349	7,016	-0,333	1,575	1,378	-0,197	6,072	5,066	-1,006	2,108	1,721	-0,387	2,598	1,837	-0,761	4,852	4,304	-0,548	4,137	3,119	-1,018	4,114	3,729	-0,385	4,077	3,442	-0,635
T18	7,404	6,947	-0,457	1,593	1,347	-0,246	6,122	5,438	-0,684	2,352	1,801	-0,551	2,759	2,181	-0,578	4,867	4,404	-0,463	4,254	3,908	-0,246	4,237	3,606	-0,631	4,154	3,235	-0,826
C19	6,963	7,032	0,069	4,937	5,356	0,419	5,502	5,163	-0,339	2,141	1,914	-0,227	1,703	1,489	-0,214	4,905	4,519	-0,386	4,217	2,156	-2,061	4,122	3,306	-0,816	4,044	2,892	-1,152
G20	7,739	7,658	-0,081				5,838	5,887	0,049	2,627	2,432	-0,195	2,497	2,198	-0,299	4,885	4,516	-0,369	4,272	4,047	-0,225	4,024	3,928	-0,096	3,921	3,894	-0,027

For each deoxynucleotide proton the value in italics (column 3) represents the shift calculated from subtracting the duplex value (column1), from the adduct value (column 2).
Chemical shift differences (italics) were underlined if the difference was greater than 0.25ppm.
Table 5 The chemical shifts for the non-exchangeable protons of adozelesin.

H3	H6	CH3	H8A2	H8A	H8B	H1A	H1B	H3'1	H4'1	H6'1	H7'1	H3'2	H4'2	H5'2	H6'2	H7'2
7,901	7,031	2,896	4,300	3,730	5,153	3,715	4,813	7,450	7,755	8,110	7,623	7,903	7,822	7,475	7,308	8,000

Table 5 The chemical shifts of all non-exchangeable protons in the DNA duplex and adozelesin-DNA adduct.

The cross-peak H8B→H8A2 was slightly less intense, whilst H1B→H8A2 seemed to be obscured by the water suppression in the TOCSY spectrum, although this correlation produced a large NOE in the NOESY spectrum.

4.5.6 Assignment of NOE connectivities between DNA and adozelesin

Once the non-exchangeable protons of both adozelesin and DNA had been assigned comprehensively, the two-dimensional NOESY spectrum was examined for NOE connectivities between protons of adozelesin and 5′d($C^1G^2A^3A^4A^5A^6A^7C^8G^9G^{10}$)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA. In all, 71 DNA→adozelesin connectivities were found, confirming the location of the subunits (A,B,C) of adozelesin within the minor groove. These NOE connectivities are illustrated in Figure 74. Additionally there were 25 intradrug NOEs between the protons of adozelesin, plus 534 NOEs between DNA base→base, base→sugar, sugar→sugar proton correlations. This led to a grand total of 630 NOE cross-peaks for distance-range constraint generation, followed by application to the DNA-adozelesin adduct molecular model.

4.5.7 The 600MHz ROESY spectrum

The ROESY spectrum for the adozelesin-5′d(CGAAAAA*CGG)·5′d(CCGTTTTTCG) DNA adduct appears in Appendix XVII. Only the most intense cross-peaks seen in the NOESY spectrum can be identified in the ROESY. ROESY data can provide information on molecular movement within a complex *e.g.* Hoogstein base-pairing or base-pair opening (Seaman and Hurley, 1996). This data set did not reveal any evidence for a conformational change of this nature. Therefore, after adozelesin has bound to the minor groove of this DNA duplex, the helix retained its β-helical character with Watson-Crick base-pairing.

4.6 Refined molecular models of the adozelesin-5′d(CGAAAAA*CGG)·5′d(CCGTTTTTCG) adduct

After the NOESY, DQF-COSY, TOCSY and ROESY data had been completely assigned, distance-range contraints were generated and applied to a computer molecular model during mechanics and dynamics calculations to yield a refined molecular structure for the adozelesin-DNA adduct. This was studied to ascertain exactly how adozelesin bound to 'A' tract DNA.

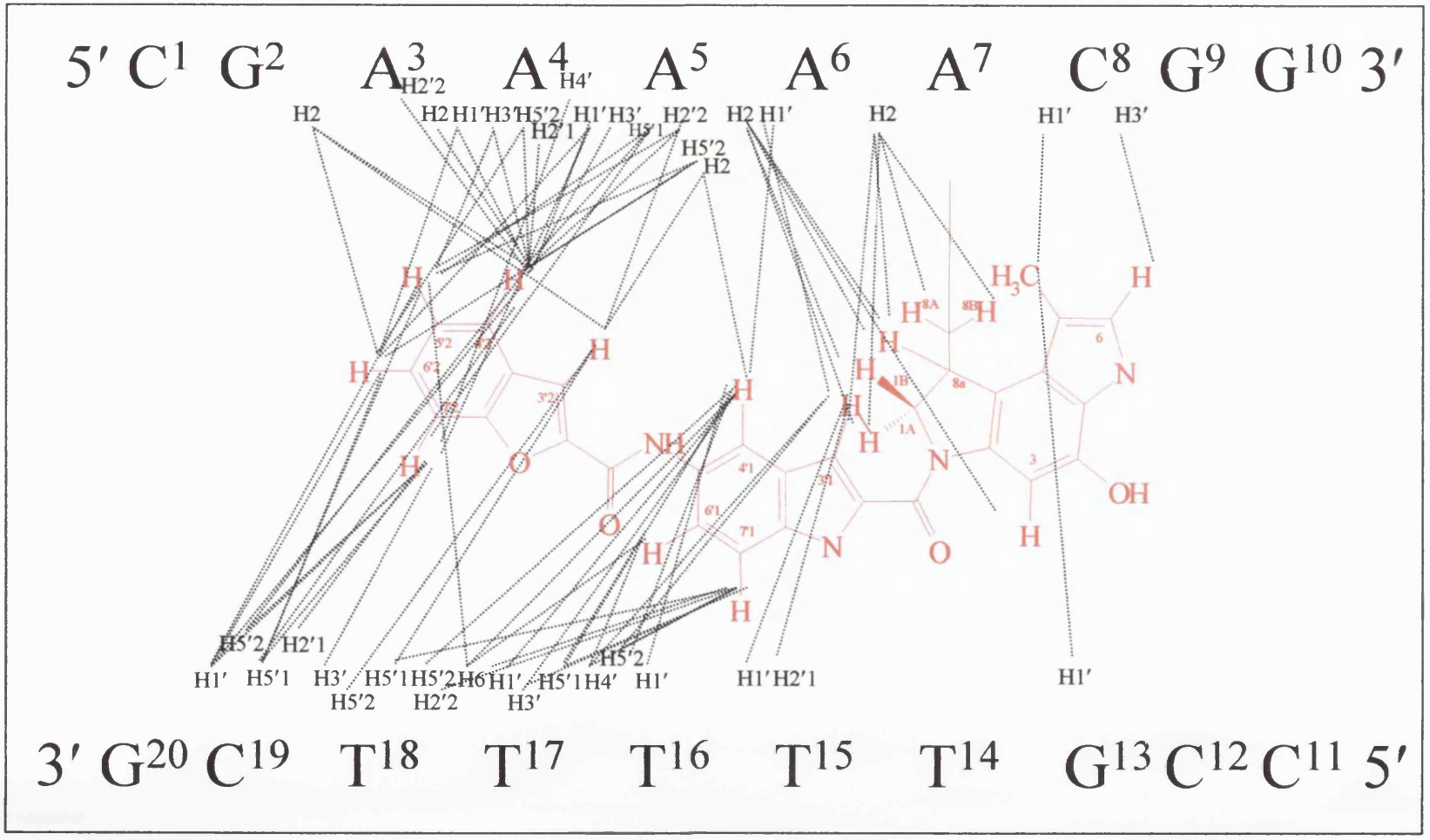


Figure 74 NOE connectivities between 'A' tract DNA and adozelesin.

The peak-pick spreadsheet (Appendix XVIII) contains a complete list of all the DNA-DNA, adozelesin-adozelesin, DNA-adozelesin NOE cross-peaks identified in the 600MHz NOESY spectrum. This was generated using the master spreadsheet (Appendix XVIV). Columns 6 and 7 respresent the lower and upper bounds (in Å) for the distance-range constraints, which were applied to the adozelesin-DNA adduct molecular model within SYBYL. Molecular mechanics and dynamics calculations were performed on this restrained molecular model both *in vacuo* and *in aquo* (when the system was solvated within a water droplet). The NOESY spectrum itself is accumulated as the DNA-adozelesin adduct, rapidly interconverts between local conformations on the NMR timescale. This internal molecular motion affects nuclear spin-spin coupling, especially NOE intensities and chemical shift. Average values of these parameters will therefore be manifested in the NOESY spectrum (Karplus and Petsko, 1990). An average structure was calculated from the 50ps dynamics trajectory, where the generated structures correspond to low energy conformations around a global minimum. Taking the mean of the dynamics data produced a structurally refined molecular model of the adozelesin-DNA adduct to fit the average conformation defined by the NMR data most closely.

4.6.1 The refined molecular model of the adozelesin-5′d(CGAAAAA*CGG)·5′d(CCGTTTTTCG) adduct generated *in vacuo*

Figure 75 The refined molecular model of the adozelesin-5′d(CGAAAAACGG)·5′d(CCGTTTTTCG) DNA adduct, generated *in vacuo*.

Figure 75 shows the *in vacuo* computer molecular model of the adozelesin-5′d($C^1G^2A^3A^4A^5A^6A^{7*}C^8G^9G^{10}$)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct, yielded after the refined molecular mechanics and dynamics calculations. This structure correlated closely with the two-dimensional NOESY data from which it was generated, with β-helical structure and Watson-Crick base-pairing being maintained. Adozelesin in Figure 75 is represented by the space-fill rendering, whilst the DNA is in stick-form. Blue denotes the 5′C^1 strand whilst magenta portrays the 5′C^{11} DNA oligomer. Examination of this model makes it easy to visualize the close-complementarity of the fit between adozelesin and its binding site in the minor groove of DNA. The three subunits of adozelesin traverse six base pairs between $C^8 \rightarrow A^3$, yet no distortion of the 'A' tract is apparent either from the NMR or modelling data, adozelesin being able to fit neatly within the minor groove. The top and bottom edges of the CPI and benzofuran rings are situated at each end of the drug binding site where the helix can be seen to turn the open cavity of the minor groove to face the back of the picture. From this point, it can be seen that they do not protrude outside the minor groove, the edge of the CPI perhaps filling the entire depth of the minor groove so that it is level with the sugar-phosphate backbone which constitutes the wall of the minor groove.

4.6.2 The refined molecular model of the adozelesin-5′d(CGAAAAA*CGG)·5′d(CCGTTTTTCG) adduct generated *in aquo*

In vacuo models are not a true representation of a biological system where for example, DNA in a cell would be surrounded by water and ions. Water, as discussed in section 3.5, is highly polar with a capacity for hydrogen-bonding. It influences greatly the macromolecular tertiary structures which it solvates, helping to maintain an equilibrium between noncovalent electrostatic, hydrogen- and van der Waals forces within the macromolecule *e.g.* β-helix of DNA (Westhof *et al.*, 1995). Therefore, explicit treatment of the solvent is necessary to produce an accurately refined molecular model, representing the adozelesin-DNA adduct in solution. 4000 Gasteiger-Hückel charged water molecules were used to solvate this adduct in a droplet. After equilibration and restrained molecular mechanics and dynamics calculations, an average structure was produced from the 501 low energy conformations of the molecular dynamics trajectory.

Figure 76 displays a three-dimensional stereoview of the adozelesin-5′d(CGAAAAACGG)·5′d(CCGTTTTTCG) DNA adduct generated *in aquo*. The water molecules are not shown, so that a clear view of the adduct is available. Once again, β-helical structure with Watson-Crick base pairing has been maintained, as indicated by the NMR data. This space-filled model correlates closely with the NOESY data. Blue atoms represent

nucleotides $C^1 \rightarrow G^{10}$, whilst magenta portrays the $C^{11} \rightarrow G^{20}$ oligomer and yellow depicts adozelesin bound within the minor groove. Examination of this 3D stereoview demonstrates the high level of complementarity achieved by adozelesin on binding to DNA within the minor groove. Adozelesin is immersed within the groove, with only its hydrophilic spine exposed to the surrounding solvent. The concave hydrophobic edge appears to be following the floor of the minor groove, which will allow the subunits of adozelesin to maximize noncovalent interactions with the minor groove. The planes of the adozelesin subunits also show different degrees of twist relative to one another. The CPI headunit and indole are nearly co-planar being only -1·1° out-of-plane, whilst a much larger twist occurs between the indole and the benzofuran (+21·4°). This twist allows the third subunit to follow the curvature of the helix more comfortably, without the drug molecule protruding out of the groove which would be energetically unfavourable. Twisting of the adozelesin molecule increases the complementarity of the fit between the drug and its binding site. Isohelical binding of drugs within the β-helix has been described for many other classic crescent-shaped minor groove binders; the cyclopropapyrroloindoles *e.g.* (+)CC-1065 (Chidester *et al.*, 1981), the oligopeptide antibiotics *e.g.* netropsin and distamycin (Berman *et al.*, 1979; Gurskaya *et al.*, 1979), the bis-benzimidazoles *e.g.* Hoechst 33258 (Searle and Embrey, 1990) and the bisamidines *e.g.* berenil (Pearl *et al.*, 1987).

Analysis of each dynamics trajectory of both the *in vacuo* and *in aquo* simulations, demonstrated how the binding of adozelesin within the minor groove of DNA behaved like a third strand encased by each DNA oligomer. This resulted in stiffening the helix, reducing its mobility around the binding site. Helix stiffening by adozelesin was also noted in the self-complementary adozelesin-DNA adduct (section 3.7) and has been previously described by Sun and Hurley, (1992b). Adozelesin did not produce winding of the 'A' tract helix like (+)CC-1065 (Lee *et al.*, 1991), with 10 base pairs still constituting a turn of the 'A' tract adduct helix.

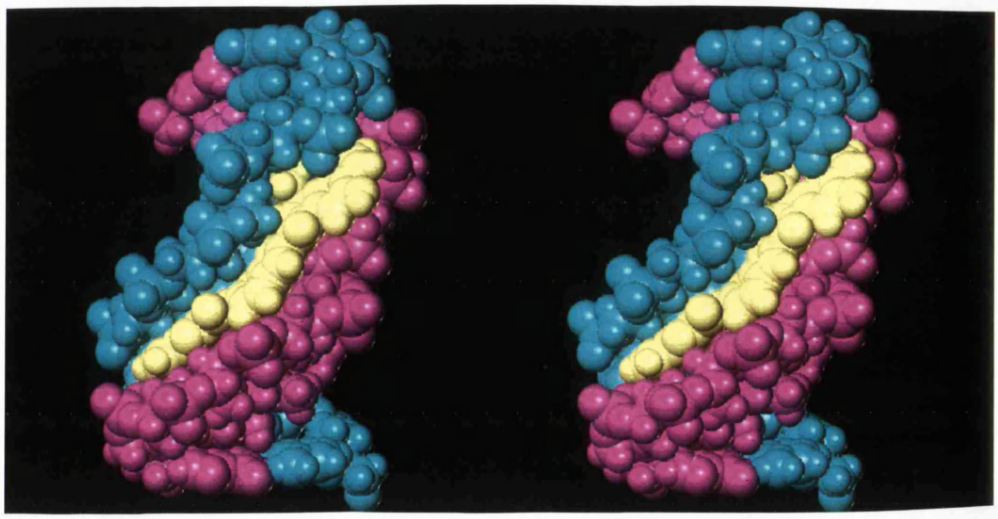


Figure 76 Stereoview of the refined molecular model of adozelesin-5′d(CGAAAAACGG)·5′d(CCGTTTTTCG) DNA adduct, generated *in aquo*.

4.7 Adozelesin covalently binds to the minor groove of DNA *via* the N3 of adenine

Alkylation of the 5′d($C^1G^2A^3A^4A^5A^6A^7C^8G^9G^{10}$)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA duplex by adozelesin, was hypothesized to take place at A^{7*}. NOE connectivities from the 600MHz NOESY spectrum were able to confirm the covalent modification site as the N3 of adenine (A^7). Figure 58 shows the mechanism of adozelesin binding to adenine. Downfield shifts of the protons derived from the cyclopropyl ring of adozelesin *i.e.* H8A at δ 1·47ppm and H8B at δ 2·02ppm (Walker *et al.*, 1997), demonstrated that covalent bond formation with adenine was *via* the C8 methylene bridge.

A^7H2 is positioned at the base of the minor groove and shows NOE correlations to those protons on the CPI headunit which are also located at the bottom of this groove *i.e.* H8A, H8B, H1A, H1B, and H8A2. These protons are located on the concave surface of adozelesin. No NOE connectivity is observed to the aromatic H3 proton situated on the opposite side of the ring system, signifying this edge of the drug is not directed into the groove. This information implies that the CPI ring system is orientated 'edge on' within the minor groove, with the hydrophobic concave surface of adozelesin binding isohelically with the floor of the groove. The A^7H2 proton has experienced a significant deshielding effect, moving its chemical shift downfield by 0·404ppm on adduct formation. A^7H2 is obviously in the same plane as the CPI head-unit which is orientated 'edge-on' within the groove, resulting in the deshielding of this aromatic H2. Sugar protons of A^7 have also been affected by the presence of adozelesin, notably A^7H1′ and A^7H4′ whose resonances have moved upfield by -0·791ppm and -1·893ppm, respectively. The shielding effect on these protons is driven by the fact they are facing directly into the aromatic 6-membered ring of the CPI headunit. Meanwhile, C^8H1′ is shielded (-0·4ppm) because it is adjacent to the top 5-membered aromatic ring of this headunit.

Electrons circulating within a benzene ring which is subjected to a magnetic field, yield diamagnetic effects *i.e.* they generate a secondary induced magenetic field (Morrison and Boyd, 1992; Abraham *et al.*, 1988). The circulation of the π eclectrons can generate a field which either opposes or reinforces the applied field felt by the proton. For benzene protons in the same plane as the ring, the induced field reinforces the applied field, so the field felt by this proton is increased and the proton is said to be deshielded. Conversely, a proton facing the ring system would be aligned so that it is subjected to a field which opposes the applied field, so the field this proton feels would be decreased and the proton is said to be shielded. Shielding leads to upfield shifts (δ) of protons whilst deshielding leads to downfield shifts of protons. These upfield and downfield shifts of DNA and adozelesin protons are apparent in Table 5 (columns in italics). The negative value depicts an upfield shift (shielding) of the proton on adduct formation, whilst a positive value indicates the proton is deshielded (downfield shift).

$G^{13}H1'$ and $C^{8}H1'$ show NOE correlations with the methyl group of adozelesin, suggesting that it too is located deep in the groove because this is where the H1′ deoxyribose protons are situated. The fact that both the cytosine and guanine yielded an NOE to this methyl indicates that the CPI ring of adozelesin is lying 'edge-on' in the centre of the minor groove. The adjacent base pair to the covalent modification site in the 5′-direction, *i.e.* $A^{6} = T^{15}$, displays NOE correlations to the protons of the pyrrolo ring of the CPI headunit. Firstly, cross-peaks between $A^{6}H2$ (situated in the centre of the base of the groove) and H1A, H1B, H8A2 and H8A, reveal that the pyrrolo ring of the CPI subunit is positioned towards the $A^{6} = T^{15}$ base pair. This large CPI ring system therefore spans just over two base pairs with its tip projecting towards $G^{9} = C^{12}$, whilst the pyrrolo end is directed towards $A^{6} = T^{15}$. Secondly, not only are the H1A, H1B, H8A2 and H8A protons situated complementary to the base of the minor groove, symbolizing 'edge on' binding, but these are NOE connectivities from sugar protons low on the wall of each side of the groove to H1A. This agrees with the set of NOEs from $C^{8}H1'$ and $G^{13}H1'$ to the methyl of adozelesin (~0·4Å), which indicated that the ring system was lying 'edge-on' centrally within the minor groove, equidistant from each oligonucleotide strand. Figure 77 shows the cyclopropapyrroloindole headunit covalently bound to the N3 of A^{7}.

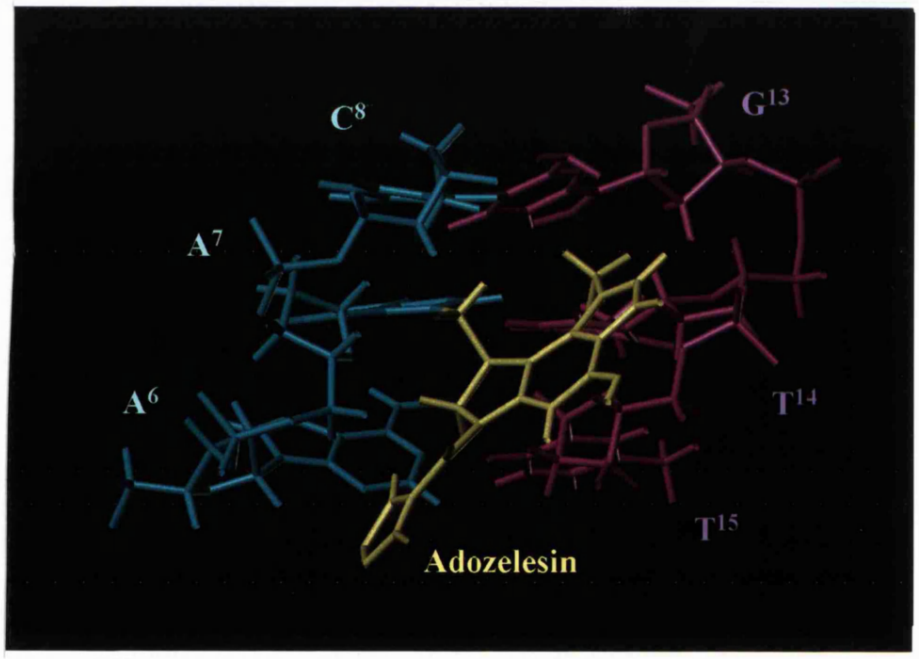


Figure 77 The cyclopropapyrroloindole headunit of adozelesin covalently binds to $A^{7}N3$ in an 'edge-on' orientation, centrally within the minor groove.

4.8 Noncovalent binding of adozelesin

An abundance of adozelesin-5′d($C^1G^2A^3A^4A^5A^6A^{7*}C^8G^9G^{10}$)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) interproton connectivities were observed in the 600MHz NOESY spectrum for defining the location of the indole (B) and benzofuran (C) subunits. NOEs from nucleotides T^{15}, T^{16} and T^{17} to the indole subunit suggest that it spans just over two base pairs. NOEs from A^3, A^4 and A^5 on one strand plus C^{19} and T^{18} on the unmodified strand to the benzofuran subunit suggest it is situated quite centrally within the minor groove, traversing just over two base-pairs. In contrast to the adozelesin-5′d(CGTAAGCGCTTA*CG)$_2$ DNA adduct discussed in chapter 3, which spans approximately five base-pairs, adozelesin in this adduct spans nearly to six. Firstly, this is interpreted from the NOEs occurring furthest in the 3′-direction along the covalently modified strand *i.e.* across $C^8 \equiv G^{13}$, where the tip of the headunit is situated. Secondly, NOE connectivities between A^3 and C^{19} to the benzofuran ring proton H6′1, on the tip of this third subunit, place the benzofuran between these two sets of base pairs $A^3 = T^{18}$ and $G^2 \equiv C^{19}$.

4.8.1 The indole subunit

Both A^6H2 and A^5H2 show NOEs to the indole protons on the second subunit. A^6H2 also displays correlations to the protons on the CPI headunit, suggesting that its position could be more adjacent to the linker half way between each ring system. As already described for the CPI headunit, the fact that these H2 protons only exhibit correlations with H3′1 and H4′1 but not with H6′1 or H7′1 indicates that this ring is inserted deeply into the minor groove in an 'edge-on' orientation. The hydrophobic, concave surface of adozelesin faces the floor of the minor groove to maximize the complementary fit of adozelesin with the β-helix. Weak NOEs from both strands, *i.e.* A^6H1′, T^{15}H1′, T^{16}H1′, T^{17}H1′ and T^{15}H2′1 which are positioned low on the wall of the groove, to H3′ and H4′1 imply that the indole like the CPI, is situated centrally and 'edge-on' within the minor groove.

Other NOE connectivities to the indole subunit originate mainly from the nucleotide T^{17}, which shows correlations the six membered ring. Measuring the interproton distances between T^{17}H1′→H6′1(2·87Å), H7′1(4·22Å) and H4′1(3·39Å) supplies the reason for the upfield shift of these sugar protons (-1·0ppm). T^{17}H1′ is facing directly into the benzene ring of the indole subunit where it experiences shielding. The T^{17}H2′1, T^{17}H2′2 and T^{17}H3′ also feel this shielding effect, though to a lesser extent. Similarly, T^{17}H4′ displays a large upfield shift from the shielding it experiences from the indole. Like H5′1 and H5′2, H4′ protons are located high up on the walls of the minor groove. Hence, strong NOE correlations to H6′1(3·92Å) and H7′1(2·48Å) are observed, proving that the convex hydrophilic surface of adozelesin is exposed

to the solvent environment outside the minor groove. Because no NOEs were detected from *e.g.* A^5H4', H5′1, H5′2 to H6′1 or H7′1, the data seem to suggest that the convex edge of the indole subunit is tipped towards the noncovalently modified strand. This would be in keeping with the co-planarity of the first two subunits, where the end of the indole ring is unable to follow perfectly the curve of the groove, so it veers closer to T^{17}. In order to maintain isohelical binding the third subunit corrects this deviation by twisting 21·4° as described earlier. This results in the central alignment of the benzofuran and hence, an improved complementary fit of adozelesin within the minor groove.

4.8.2 The benzofuran subunit

A grand total of 41 NOE connectivities between the terminal benzofuran subunit and nucleotides A^3, A^4, A^5, T^{18} and C^{19} were identified in the NOESY spectrum, of which 16 NOEs were attributable to the noncovalently modified strand and 25 NOEs to the alkylated oligomer. On initial inspection, it is easy to assume that this subunit is therefore more closely associated with the modified strand. However, some of these NOEs are to adenine H2s which are located in the centre of the minor groove. Adenines A^5H2, A^4H2 and A^3H2 each show correlations to H3′2, H4′2 or H5′2, with no NOE to H7′2, confirming that the benzofuran, like the first two subunits, is orientated 'edge-on' within the minor groove. A^5H2 shows a large downfield shift and inspection of the molecular model (Fig. 78) shows, it is directly in plane with H4′1(2·3Å) and H3′2(2·7Å) although it is adjacent to the linker between the indole and the benzofuran. The downfield shift is therefore a result of the deshielding effects A^5H2 experiences from the B and C subunits. Owing to the H1′, H2′1, H2′2 and H3′ protons situation low down on the wall of the minor groove, correlations from these protons to H3′2, H4′2 and H5′2 also testify to the 'edge-on' orientation of the benzofuran.

The H4′, H5′1 and H5′2 protons are located high on the walls of the minor groove. NOEs from these protons were generally to H6′2 and H7′2, indicating that H6′2 and H7′2 were directed outwards from the base of the minor groove. This is in agreement with the convex, hydrophilic surface of the benzofuran being able to interact with the solvent surrounding the adduct. C^{19} shows some huge changes in chemical shift of its H4′, H5′1 and H5′2 sugar protons even though it is five base-pairs away from the alkylation site and on the opposite strand. Analysis of the refined molecular models show that the large shielding effect is due to the closeness of H5′2(3·81Å), H6′2(2·53Å) and H7′2(2·32Å). A^4H4' and A^5H4' have also experienced upfield shifts, resulting from the shielding of the 6- and 5-membered rings of the benzofuran respectively. The centralized, 'edge-on' orientation of the benzofuran maximizes the complementarity observed between drug and DNA, to produce a perfect isohelical fit of adozelesin within the 'A' tract of β-helical DNA(Fig. 78).

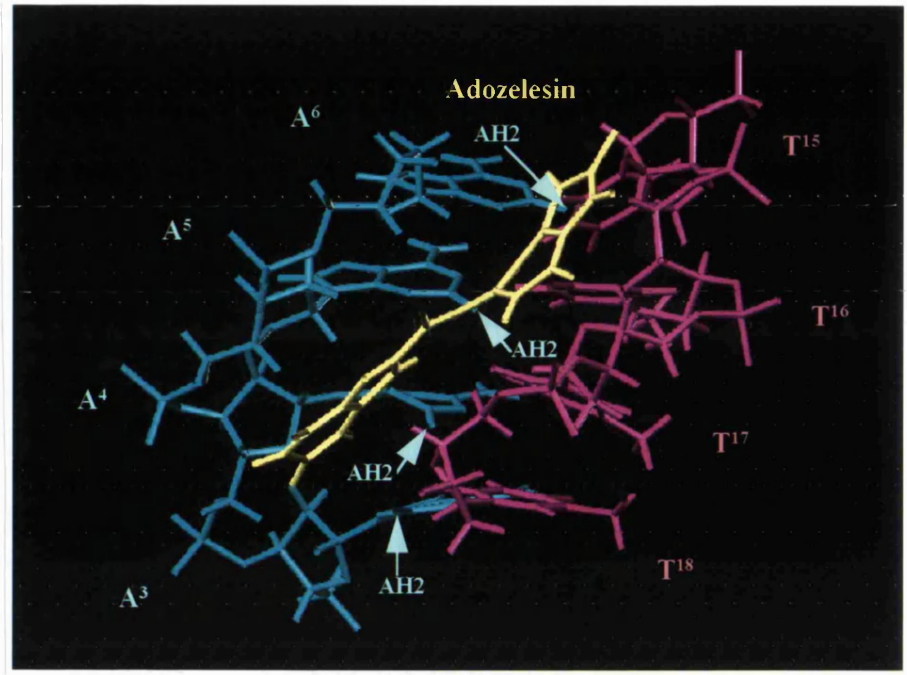


Figure 78 The indole and benzofuran subunits of adozelesin bind noncovalently into the minor groove of DNA.

4.9 Adozelesin displays sequence selectivity towards an adenine-rich 'A tract' DNA minor groove

Analysis of the NOESY, DQF-COSY and TOCSY NMR data sets have shown that there is only one fully assignable major adozelesin-DNA species present within this sample. All the NOE connectivity 'walks' are present and each proton (except those related to G^{10}) is accounted for. The DNA duplex 5′d($C^1G^2A^3A^4A^5A^6A^7C^8G^9G^{10}$)· 5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) was non-self-complementary and hence, so was the adozelesin-DNA adduct, each strand having its own separate set of sequential cross-connectivities. Since adozelesin is known to traverse 5bp in the adozelesin-5′d(CGTAAGCGCTTA*CG)$_2$ DNA adduct (see Chapter 3 and Cameron and Thompson, 1999), there is only space in this non-palindromic DNA helix for one adozelesin to bind. Covalent binding is most likely to take place at A^7 because this is the adenine at the 3′-end of the 'A' tract. Hurley *et al.*, (1984) decribed 5′d(AAAAA*) as a consensus sequence for (+)CC-1065. Reynolds *et al.*, (1985) found that the 2nd and 3rd adenines from the 3′ end were more well

conserved than the 4th and 5th adenines in the sequence, which the third benzodipyrrole subunit of (+)CC-1065 interacted with. This evidence suggests a run of five sequential adenine bases is the most preferred binding site for (+)CC-1065 binding and it would follow that adozelesin would react in a similar manner. Boger *et al.*, (1991b) noted that the 'A' tract sequence was preferred over all other binding sites, for a variety of CPI analogues. This observation was confirmed specifically for adozelesin by Weiland and Dooley, (1991). A^6 would perhaps be the second best candidate for covalent modification and may be the source of the minor species detected in the DQF-COSY and TOCSY spectra, as previously discussed.

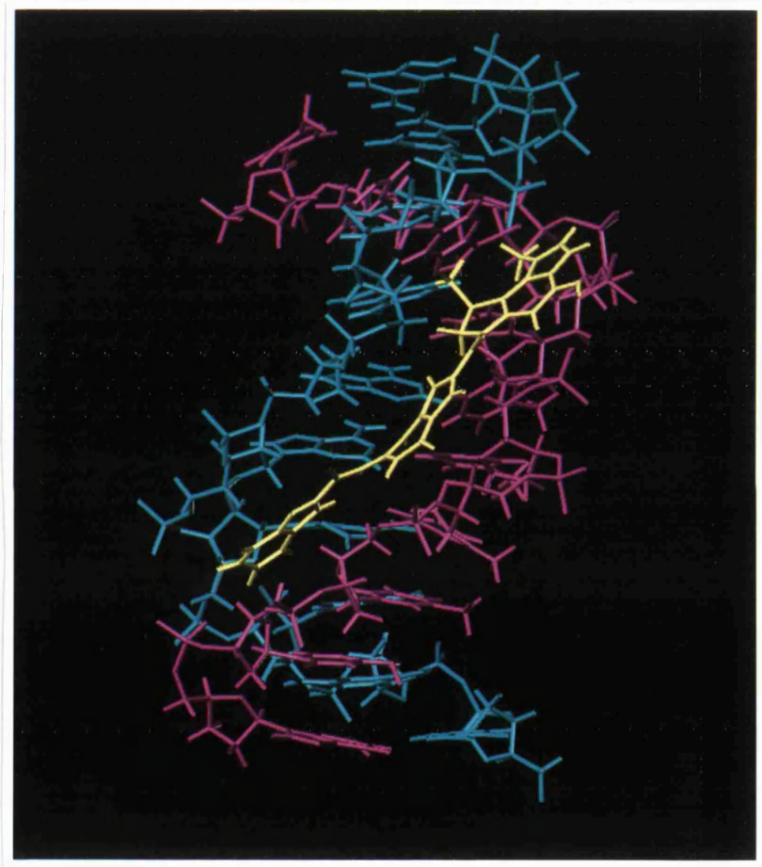


Figure 79 The DNA duplex 5′d($C^1G^2A^3A^4A^5A^6A^7C^8G^9G^{10}$)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) retains 'A' tract character on covalent modification by adozelesin.

The indole (B) and benzofuran (C) subunits of adozelesin are responsible for the noncovalent binding mode of this drug. Noncovalent binding is known to be the source of sequence selectivity for a drug binding to DNA, irrespective of whether covalent modification is involved. It is this noncovalent binding mode which dictates the flanking sequence around the binding site, yielding a specific binding sequence so that adozelesin does not indiscriminately alkylate every adenine in DNA. Hurley and Needham-VanDevanter (1983) attributed the

binding specificity of (+)CC-1065 to the hydrophobic interactions between its concave surface and the adenine H2 protons. The van der Waals contacts between the ligand and the narrow groove maximize the affinity of the ligand, by producing a strong force of attraction at a specific binding sequence (Boger *et al.*, 1991b). These noncovalent interactions will be particularly prevalent in the narrow 'A' tract minor groove. It is likely that adozelesin preferentially alkylates A^7 over A^6 (possible minor species) for this reason, allowing the entire adozelesin molecule to be submerged in the very narrow 'A' tract minor groove for maximization of noncovalent binding.

Inspection of both the NOESY data and the resulting molecular model demonstrates that this DNA sequence has retained its 'A' tract character. For example, many protons of C^{19} display NOEs to adozelesin, whereas, in the self-complementary adduct, the equivalent nucleotide G^8 on the unmodified strand was not involved in any NOE connectivities to the benzofuran 'tail'. The NOEs to the benzofuran from C^{19} would not be possible without propellor twisting of bases. All adenines displayed propellor twisting and, although Watson-Crick base pairing is maintained, bifurcating hydrogen-bonds are clearly present between the N6 amino of adenine and O4 of the thymine, constituting the AT base pair to the 3′-side of adenine (Nelson *et al.*, 1987; Katahira *et al.*, 1988). The propellor twisting of bases narrows the minor groove, producing a more compact fit between adozelesin and its binding site within DNA. This in turn optimizes the noncovalent binding mode (Fig. 79). 'A' tract DNA has only 10bp per turn (Rhodes and Klug, 1981) and is straight (Nelson *et al.*, 1987). In this environment adozelesin binds in the centre of the minor groove, in an 'edge-on' orientation, spanning over 6bp. It is the straightness of this groove which perhaps allows adozelesin to bind so neatly, without producing tilting of the second and third subunits as observed in the adozelesin-5′d(CGTAAGCGCTTA*CG)$_2$ adduct in Chapter 3. The fact that *ca.* 6 base pairs are traversed by adozelesin instead of 5 may also reflect the tighter helical turn of an 'A' tract *i.e.* 10bp instead of 10·5bp.

In the self-complementary adozelesin-5′d($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ adduct, a distortion was noted around the base G^6 (section 3.8.5.6). No deformation of this kind has been detected in the 'A' tract adduct, although the sixth flanking base pair in the 5′ direction along the modified strand happens to be C^1, a terminal nucleotide. The H1′, H2′1 and H2′2 all display unusually high chemical shifts which there seems to be no reasonable explanation for. Terminal nucleotides would be expected to produce very similar chemical shifts in duplex and adduct DNA, because they are not part of the binding site. However, if adozelesin does cause a distortional wave 6bp from the binding site on the modified strand, as indicated by Seaman *et al.* (1996) and Lin *et al.* (see review by Hurley and Draves, 1993), then this distortion would

culminate at C^1, hence explaining the strange chemical shifts. A longer oligonucleotide would be required to investigate this further.

4.10 Hydrogen-bonding

Analysis of the binding modes of minor groove binding agents, such as the oligopeptide antibiotics (Zakrzeqska *et al.*, 1983), the bis-benzimidazoles (Parkinson *et al.*, 1994), the bis(quarternary ammonium) heterocycles (Gao *et al.*, 1993) and the cyclopropapyrroloindoles (Hurley and Needham-VanDevanter, 1986), has demonstrated that hydrogen bonding only contributes to the stability of a complex.

Unlike the adozelesin-5′d(CGTAAGCGCTTA*CG) DNA adduct, this 'A' tract adduct provided no evidence of hydrogen-bonding between C^8O1P phosphate anionic oxygen and the CPI hydroxyl. The central location of the headunit means that the closest moiety with which it could hydrogen bond is T^{16}O1P, but this is 3·95Å away, nearly twice the length of the H-bond. No water molecules were seen to be bridging a bond such as this during the dynamics calculations (Lin *et al.*, 1991). The only possible hydrogen-bond is from the NH of the amide linker, which is between the indole and benzofuran subunits of adozelesin to the N3 of adenine A^5. This distance measured in the refined molecular model is 2·830Å, which is rather long for a hydrogen-bond (usually ~2Å). However, the close complementarity of the fit of adozelesin within the narrow 'A' tract minor groove is likely to produce an abundance of other noncovalent interactions, which will more than compensate for the lack of stabilizing hydrogen-bonds.

4.11 In Summary

The cyclopropapyrroloindole adozelesin binds to the minor groove of the DNA sequence 5′d($C^1G^2A^3A^4A^5A^6A^{7*}C^8G^9G^{10}$)·5′d($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$). Alkylation took place on cyclopropyl ring-opening, to the N3 of adenine (A^7) of the consensus sequence 5′d($A^3A^4A^5A^6A^{7*}$). Adozelesin bound with its benzofuran 'tail' orientated towards the 5′-terminal of the modified strand. This adduct was non-self-complementary in nature. The adenine base H2 protons (A^7H2 and A^6H2) were used to define the position of the CPI (A) headunit within the minor groove. Adozelesin was found to bind 'edge-on', along the centre of the minor groove. The convex, hydrophilic surface was directed out of the groove, while the concave, hydrophobic surface lay along the floor of this groove. The crescent shape of adozelesin resulted in this ligand binding isohelically within the β-helical DNA.

The benzofuran and indole subunits of adozelesin controlled the sequence selectivity element of binding. This sequence selectivity is a result of noncovalent interactions such as hydrophobic interactions and van der Waals forces between adozelesin and DNA. In all adozelesin spanned over 6 base pairs, with the 'A' tract retaining its propellor twisting conformation on adduct formation. The conformation of the flanking sequence and the topology of adozelesin is extremely important for allowing adozelesin, to wedge itself tightly into the 'A' tract minor groove of DNA in a sequence selective manner.

The preference displayed by adozelesin towards these 'A' tract sequences is obviously a result of the perfect isohelical fit, with which adozelesin is able to bind within the minor groove of macroscopically bent DNA. Although there is no data on a (+)CC-1065-'A' tract DNA adduct it is likely that through the similarity of their molecular structures, the resulting adduct which displays close complementarity between ligand and binding site, will be very similar. Adozelesin however, does not display any of the winding effects observed to be present with (+)CC-1065 (Warpehoski *et al.*., 1988). The narrowness of the 'A' tract groove is conducive to the presence of noncovalent interactions with the drug. This, in conjunction with, the bend of the DNA sequence which allows adozelesin to lie 'edge-on', in a central position along the length of the groove, demonstrates why 'A' tract character easily promotes sequence selective binding of adozelesin.

CHAPTER 5: RESULTS AND DISCUSSION

EVALUATION OF THE MINOR GROOVE BINDERS; CYANOMORPHOLINOADRIAMYCIN, DSB-120 AND SIBIROMYCIN

The ability of various other minor groove binding agents to modify DNA has also been assessed by 1D ^{1}H NMR. Cyanomorpholinoadriamycin (CMA) is an anthracycline which displays a mixed binding mode. Both intercalating and covalent minor groove binding moieties are incorporated into the structure of CMA. DSB-120 and sibiromycin belong to the pyrrolo[2,1-*c*][1,4]benzodiazepine family, which covalently bind within the minor groove.

5.1 The modification of DNA by cyanomorpholinoadriamycin

Cyanomorpholinoadriamycin was allowed to react with two duplex DNA sequences; 5′d(G^{1}T^{2}T^{3}C^{4}C^{5}A^{6}T^{7}G^{8}C^{9}A^{10}A^{11}C^{12})$_2$3′ and 5′d(G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}T^{7}A^{8}I^{9}C^{10})$_2$3′. Both sequences are self-complementary, although only the former was hypothesized to form a self-complementary adduct.

5.1.1 The 1D ^{1}H NMR spectrum of 5′d(GTTCCATGGAAC)$_2$

A 1D ^{1}H NMR spectrum (Fig. 80) was collected on the purified 5′d(G^{1}T^{2}T^{3}C^{4}C^{5}A^{6}T^{7}G^{8}G^{9}A^{10}A^{11}C^{12})$_2$ DNA sample. Analysis of the non-exchangeable protons within the DNA duplex was performed as previously described in section 3.1. The palindromic nature of this duplex means both strands are equivalent, so the spectrum effectively corresponds to one DNA strand.

To begin with, the methyl region of the spectrum was examined to assess whether signals from all the thymines were present. Three methyl peaks were noted at approximately δ 1·25, δ 1·35, and δ 1·55ppm corresponding to T^2, T^3, and T^7, although not necessarily in this order. There are six A/GH8 and six T/CH6 base aromatic protons in this sequence. Thirteen resonances are immediately obvious, which included signals from the 12 H6/H8 protons, two of

which are doublets at ~δ 7·1 and δ 7·45ppm corresponding to cytosine H6. The remaining resonance corresponds to an adenine H2 proton, of which their are three in this duplex at A^6, A^{10} and A^{11}. Therefore, two resonances must be overlaid with two other peaks. The broadness of the peak at δ 7·39ppm suggests that this is the case. Although it is not possible to attribute any peak in the aromatic region of the spectrum to a specific nucleotide, the tall and needle like appearance of the peaks at ~δ 7·26 and δ 7·70ppm suggest they could be due to AH2 protons. This spectrum indicated that the DNA sample was pure, so it was allowed to react with CMA.

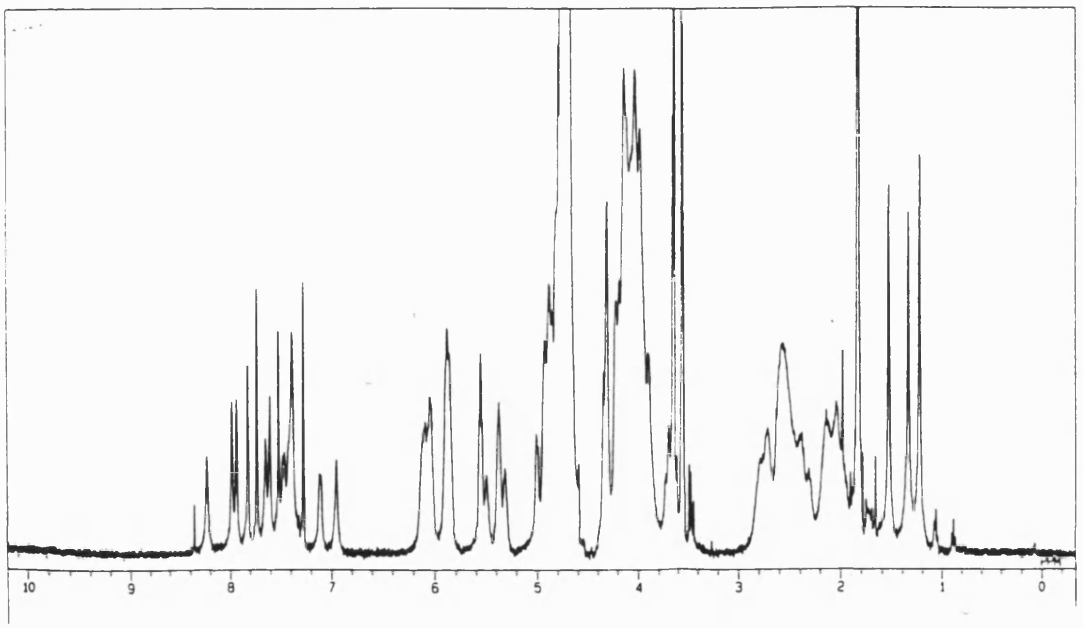


Figure 80 The 1D ^{1}H NMR spectrum of 5′(GTTCCATGGAAC)$_2$.

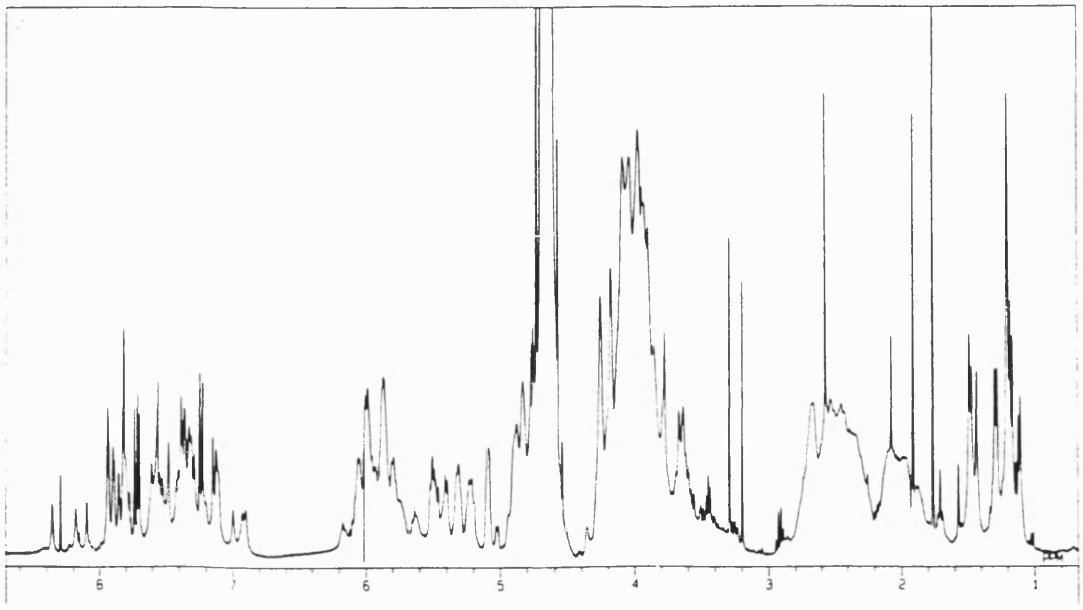


Figure 81 The 1D ^{1}H NMR spectrum of the CMA-5′(GTTCCATGGAAC)$_2$ complex.

5.1.1.2 The 1D ^{1}H NMR spectrum of the CMA-5′d(GTTCCATGGAAC)$_2$ complex

Subsequent to purification of the drug-DNA reaction mixture using C_{18} Sep Pak cartridges, a one-dimensional proton NMR spectrum was accumulated on the CMA-5′d(GTTCCATGGAAC)$_2$ complex. Inspection of these data shows that the number of peaks in this spectrum has increased significantly, indicating that a reaction has taken place.

The structure of cyanomorpholinoadriamycin is given in Figure 82. This drug has three aromatic protons corresponding to positions 1,2 and 3 on the D ring of the aglycone chromophore, whose chemical shifts would be expected to be in the region δ 7-8ppm.

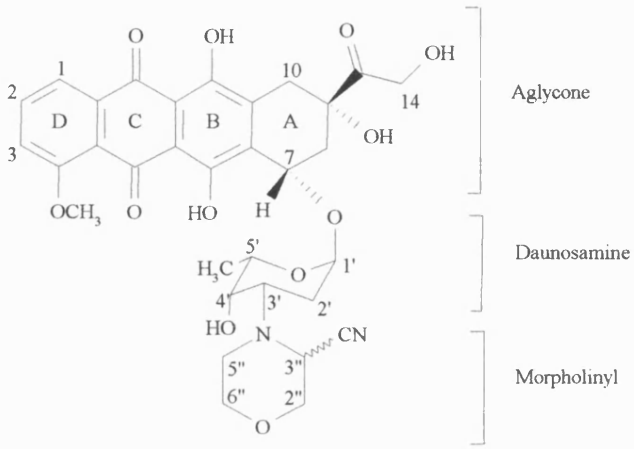


Figure 82 The molecular structure of cyanomorpholinoadriamycin (CMA).

The methyl of the daunosamine sugar and the methoxy group (which is also situated on the D ring of the aglycone), serve as the remaining two clear molecular markers for evidence of reaction between CMA and the DNA duplex. The chemical shifts of the methyl and methoxy groups were anticipated to be between δ 1·1·5ppm and δ 3·5-4·0ppm respectively.

Figure 81 shows the ^{1}H NMR spectrum of the CMA-5′d(G^{1}T^{2}T^{3}C^{4}C^{5}A^{6}T^{7}G^{8}G^{9}A^{10}A^{11}C^{12})$_2$ complex. Analysis of the methyl region should have yielded four peaks, if this palindromic sequence had produced a self-complementary adduct on modification by CMA. These would correspond to T^2, T^3, T^7 and the daunosamine 5′-methyl. Instead, at least eleven peaks existed in this methyl region. Clearly, the methyl peaks at δ 1·25, δ 1·35 and δ 1·55ppm were equivalent to the T^2, T^3 and T^7 nucleotides of unreacted duplex DNA. The intensity of these peaks suggested that only *ca.* 50% of the DNA had formed a complex with CMA, whilst 50% remained as DNA duplex. No methyl peak was expected to represent unreacted CMA because all free drug was removed during purification. This left eight

singlets at approximately δ 1·48, δ 1·30, δ 1·27, δ 1·2 (overlaid), δ 1·17, δ 1·16, δ 1·12 and δ 1·1ppm. Dividing this number by two results in the correct number of methyls for two CMA-5′d($G^1T^2T^3C^4C^5A^6T^7G^8C^9A^{10}A^{11}C^{12}$)$_2$ complexes; 2 x T^2, 2 x T^3, 2 x T^7 and 2 x daunosomine CH_3. These methyls are not equivalent because CMA binding at each 5′GG intercalation site is not in the same orientation and, hence, a self-complementary CMA-DNA adduct is not formed.

Examination of the aromatic region (Fig. 81) indicated that the number of peaks far exceeded the expected total of 18 (15 x H6/H8/H2 of DNA and 3 x $H^1/H^2/H^3$ of the aglycone of CMA), which would correspond to a self-complementary adduct. Some areas of this aromatic region between 6·8 and 8·4ppm are quite broad, owing to the peaks being so overlaid.

These data suggest that two CMA-5′d(GTTCCATGCAAC)$_2$ species are formed on mixing the anthracycline CMA with this DNA duplex sequence.

5.1.1.3 CMA intercalates at the 5′GG site

Jesson *et al.*, (1989) established that intercalation into DNA was important for CMA cross-linking activity. The level of cross-linking was also shown to increase as the DNA became more G-C rich. Cullinane and Phillips, (1991) utilised *in vitro* transcription assays to probe the sequence specificity of CMA complexes within DNA. Transcriptional blockages were noted prior to 5′-CC and 5′-GG which were thought to be consistent with intrastrand cross-links, whilst low level blockages at 5′-GC and 5′CG sequences were thought to reflect interstrand cross-links. Intrastrand cross-linking at the 5′GG site of the non-template strand, was found to predominate over 5′-CC, 5′-GC and 5′-CG base combinations (Cullinane and Phillips, 1992). The sequence specificity of CMA *in vivo* was found to be equivalent to that found with *in vitro* testing (Cullinane and Phillips, 1994). From this evidence, it was reasonable to expect the aglycone moiety of cyanomorpholinoadriamycin to intercalate preferentially between the 5′G^8G^9 base pairs in the 5′d($G^1T^2T^3C^4C^5A^6T^7G^8G^9A^{10}A^{11}C^{12}$)$_2$.

Without being able to map chemical shift differences between duplex and modified DNA to detect the exact positioning of CMA within the complex, it is apparent that the only preferred sequence available for reaction is at 5′-G^8G^9. This allows two CMA molecules to react with this β-helical dodecamer, each strand coding for one reaction site. It is obvious that some complex formation has taken place, as indicated by the complexity of the spectrum.

Formation of a self-complementary adduct would require each CMA to intercalate into the 5′-G^8G^9 sequences on each strand, in the same orientation. The daunosamine sugar and

morpholino moieties would therefore be projected towards the same terminal of the modified strand. For self-complementarity, both CMA molecules should be orientated so that their aglycone 'heads' are intercalated between 5′-GG. Their morpholino tails are then directed either outwards towards the terminals of the duplex, in the 3′-direction of the 5′-GG template strand, yielding a 'head to head' complex (Fig. 83). Conversely a 'tail to tail' complex is formed where the morpholino 'tails' are orientated towards the centre of the DNA, in the 5′ direction of the 5′-GG template strand (Fig. 84).

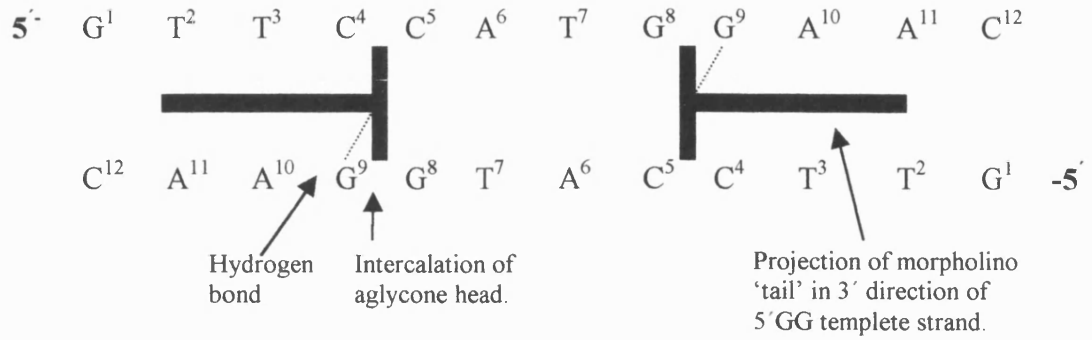


Figure 83 Schematic representation of the 'head to head' self-complementary complex formed on the reaction of CMA with DNA.

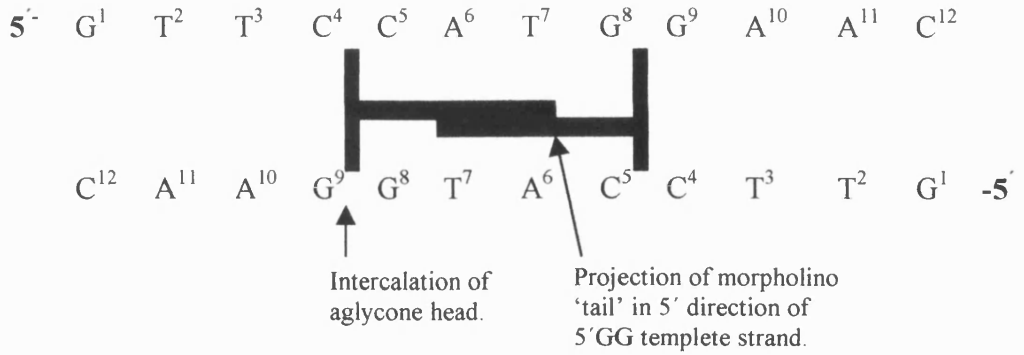


Figure 84 Schematic representation of the 'tail to tail' self-complementary complex formed on the reaction of CMA with DNA.

The 'head to head' self-complementary adduct is a probable candidate for one of the species present in the reaction mixture. T^2, T^3 and T^7 are in equivalent environments in both oligonucleotide strands, contributing three methyl resonances to the spectrum. Wang *et al.*, (1987) found that the A ring of the aglycone moiety of daunomycin projected into the minor groove. The hydroxyl group at C9 of this chromophore formed a hydrogen-bond with G^2N2 of the 5′d($C^1G^2T^3A^4C^5G^6$)$_2$ duplex, where daunomycin intercalated between the terminal 5′CG base pairs. This hydrogen-bond was actually part of a bifurcating hydrogen-bond, between the same G^2N2 and O7 of the glycosyl ether linkage. The binding of the adriamycin analogue 3′-(2-methoxy-4-morpholinyl)adriamycin to 5′d(CGTACG) was investigated by Odefey *et al.*, (1992). Hydrogen-bond detail was also described here between the aglycone O9 hydroxyl and

the N2 of guanine (G^2). Examination of Figure 83, demonstrates that hydrogen-bonding between the O9 hydroxyl of the aglycone and G^9N2 is also possible in this self-complementary CMA-DNA complex. G^9 is the 3′ guanine in the intercalation site on the template strand. The hydroxyl and aceto group lie on the same surface of the aglycone as the daunosamine sugar. However, the conformation of the cyclohexene ring dictates that C9 in not in the same plane as the aglycone, instead it is directed downwards in the same direction as the daunosamine sugar. This allows the aceto group which is on the opposite surface of the ring system, to be below the plane of the aglycone, like the hydroxyl and daunosamine. Hydrogen-bonding between the aglycone O9 hydroxyl and G^9 will anchor this edge of the chromophore into the minor groove.

Intercalation of the aglycone 'head' in the opposite orientation (Fig. 84) appears to be less favourable. Firstly, when the lower surface of the aglycone, from which the hydroxyl, aceto and daunosamine moieties project, faces in the opposite direction, the hydroxyl O9 is now closest to C^5. C^5 is in the equivalent position to G^9 (of the 'head to head' adduct) for hydrogen-bonding. The lack of a nucleophilic centre, such as the N2 of G^9, at C^5 means there will be no hydrogen-bonding to stabilize aglycone intercalation. The length of the combined daunosamine and morpholino moiety presents a second problem. For two CMA molecules to bind to 5′d(GTTCCATGGAAC)$_2$ *via* intercalation between adjacent guanines, it is very unlikely that their will be enough space to allow both 'tail' groups to fit within the minor groove. There are only two base pairs A^6 and T^7 between these intercalation sites. Odefey *et al.*, (1992) observed an NOE from the methoxy group attached to the morpholino ring, to A^4H1′ of 5′d($C^1G^2T^3A^4C^5G^6$), when intercalation took place between C^1 and G^2. This suggests that CMA spans 3-4 base pairs in total. Both morpholino tails would extend in the 5′ direction towards A^6 (on the same strand as the G^8G^9 intercalation site). However, the first CMA to bind to the minor groove is likely to obstruct the daunosamine sugar and morpholino group of the second drug molecule from binding in the same position. It is probable that the binding of CMA in this 'tail to tail' orientation would be energetically unfavourable for this reason. However, formation of a monoadduct, with the morpholino substituent of CMA projecting towards the centre of the duplex, is quite feasible. In a drug-DNA complex of this nature, all thymine methyls of the DNA would be in a different environment, resulting in six different methyl peaks in the spectrum.

The fourth species which could be present in this reaction mixture, is a 'head to tail' CMA-DNA complex (Fig. 85). Owing to the palindromic nature of the DNA, this is equivalent to a 'tail to head' description of drug binding. All thymine methyls in this CMA-modified DNA duplex will be in a different chemical environment, as will the daunosamine sugar methyls of each CMA. A total of eight separate methyl resonances would be observable from this 'head to tail' CMA-DNA complex. Although this could well represent the species seen in the one-

dimensional ^{1}H NMR spectrum, only 400 MHz two-dimensional NMR data could be accumulated on this sample.

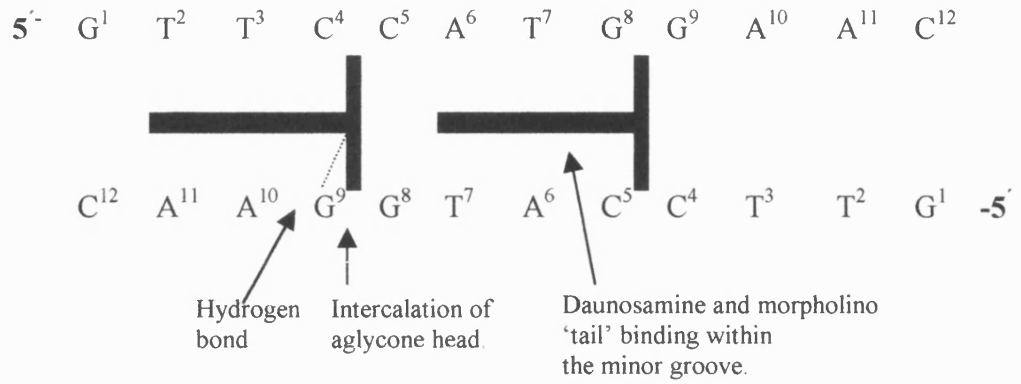


Figure 85 Schematic representation of CMA binding in a 'head to tail' orientation within the self-complementary DNA duplex.

It was thought that this data would not be of sufficiently high quality to determine the structure of the CMA-DNA adduct. As shown in Figure 85, only one aglycone is in the correct orientation to form a hydrogen bond with G^9, *i.e.* the aglycone is facing the centre of the DNA duplex. However, each CMA molecule has enough space to lay comfortably within the minor groove.

5.1.1.4 Cross-linking of DNA by cyanomorpholinoadriamycin

As previously described by Cullinane and Phillips (1991; 1992; 1994), the 5′GG site is thought to induce CMA-DNA intrastrand cross-links. Mazzini *et al.*, (1998) examined the positions of the morpholino ring substituents of 3-methoxymorpholinoadriamycin and morpholinoadriamycin, on formation of complexes with 5′d($C^1G^2A^3T^4C^5G^6$)$_2$ and 5′d($C^1G^2T^3A^4C^5G^6$)$_2$ *via* NMR. In each case intercalation took place between 5′-C^1G^2, with the aglycone hydroxyl within hydrogen-bonding distance of G^2N2. The daunosamine ring lay across G^2 and T^3 in the 5′-T^3A^4 adduct, while in the 5′-A^3T^4 sequence, daunosamine migrated closer to C^{11} and T^{10} of the opposite strand. The morpholino group exhibited NOEs to T^4 of the 5′AT complex and to A^4 of the 5′TA CMA-DNA complex. Hence, the morpholino moiety of CMA would be expected to be adjacent to T^7 in the CMA-5′d(GTTCCATGGAAC) complex, when the aglycone intercalated between G^8 and G^9, with the tail projecting towards the centre of the oligomer. Conversely, if CMA was facing in the opposite direction, the morpholino group could be hypothesized to be located next to A^{11}.

On cross-link formation, cyanide is lost from CMA to yield an iminium ion which can react with the nucleophilic centre of G-N2 (Westendorf *et al.* 1989). The cyanomorpholino

substituent in the CMA-5′d(GTTCCATGGAAC) adduct is proposed to be adjacent to either an adenine (A^{11}) or a thymine (T^7), depending on the orientation of the drug, An inappropriate design of this DNA duplex means there is no guanine in an adjacent position to the cyanomorpholino group. Therefore, no covalent bond can be formed between the cyanomorpholino moiety and G-N2 by the loss of cyanide.

CMA has bound to 5′d(GTTCCATGGAAC) *via* intercalation of the aglycone but the formation of a stabilizing hydrogen-bond from the O9 hydroxyl to G-N2 will depend on the orientation of the CMA within the minor groove. It is also very unlikely that any covalent modification of this DNA duplex has taken place. This is due to the lack of a guanine nucleotide N2 within two flanking base pairs either side of the intercalation site. It is possibly the absence of this guanine where a CMA-DNA covalent cross-link could have formed that is responsible for the inferior sequence-selectivity displayed by CMA on this DNA oligomer. Two base pairs 5′GG, are not enough to constitute a sequence selective binding site for the anthracycline, cyanomorpholinoadriamycin. Hence, on reaction of this drug with DNA, multiple species were formed.

5.1.2 The 1D ^{1}H NMR spectrum of 5′d(GCTAGCTAIC)$_2$

Analysis of the one-dimensional NMR spectrum (Fig. 86) of the DNA duplex 5′d($G^1C^2T^3A^4G^5C^6T^7A^8I^9C^{10}$)$_2$ demonstrated that this was a pure sample. The spectrum was referenced to the D_2O peak at δ 4·71ppm. The methyl and aromatic regions were examined in more detail, as previously described, to see if the correct number of nucleotides were present. The palindromic nature of the helix produced a single set of resonances equivalent to one strand of the duplex.

Only two singlets at δ 1·5 and δ 1·6ppm were observed in the methyl region, corresponding to T^3 and T^7. Eleven peaks were present in the aromatic region, three of which were doublets at δ 7·2, δ 7·3 and δ 7·4ppm which correspond to C^2, C^6 and C^{10}H6. The two peaks at δ 7·25 and δ 7·7ppm were particularly tall and sharp and, therefore, were likely to be the inosine (I)H2 and adenine H2 resonances. The remaining peaks at δ 6·95 (x2 overlaid), δ 7·6, δ 7·8, δ 7·9, δ 7·1 and δ 7·15ppm correlated with the TH6, GH8 and AH8 aromatic protons.

5.1.2.1 The 1D ^{1}H NMR spectrum of the CMA-5′d(GCTAGCTAIC)$_2$ complex

This DNA sequence was designed so that it contained a central 5′GC interstrand cross-linking site, as defined by Cullinane and Phillips (1992; 1993). It was proposed that CMA would intercalate between these base pairs and in doing so form an interstrand cross-link between the guanines. The terminal 5′GC sequences had an inosine incorporated into their complementary base pairs *i.e.* IC-3′, so that no N2 was available for covalent modification by CMA, hence yielding a more unfavourable binding site.

Figure 87, shows the spectrum of the CMA/5′d(GCTAGCTAIC)$_2$ DNA reaction mixture, after being stirred for 48h in the dark. All peaks are very broad and the sample is in need of purification. However, even though this sample has produced a poorly resolved spectrum, the general topology of the peaks appears to be the same as in Figure 86. Purification of the mixture by C_{18} Sep Pak cartridges, followed by NMR, yielded the spectrum in Figure 88. Overlaying Figures 88 and 86 verified that this sample was in fact unmodified DNA duplex. All proton chemical shifts were identical in both spectra. The 5′d(GCTAGCTAIC)$_2$ oligomer did not react with CMA to form an interstrand cross-link. CMA did not even intercalate at the central 5′GC in a noncovalent mode.

5.1.2.2 CMA does not react with 5′d(GCTAGCTAIC)$_2$

Cullinane and Phillips (1993) demonstrated that no cross-linking took place at 5′IC sites, which was why this pair of base pairs was chosen as the terminal of this DNA duplex. Interstrand cross-linking had been noted at 5′GC sequences (Cullinane and Phillips, 1991; 1992; 1994). This single site was incorporated into this DNA duplex, in the hope that on reaction with CMA, one species of monoadduct only would be formed.

In hindsight, the 5′GC combination of base pairs was the wrong choice of sequence to target CMA in the hope of forming an interstrand cross-link. Both Chaires *et al.*, (1987) and Wang *et al.*, (1995) described 5′(A/T)CG as the consensus sequence for binding of anthracyclines. The reversal of the 5′GC base pairs to 5′CG, puts the guanine N2 in a prime position for hydrogen bond formation with the CMA aglycone hydroxyl at C9 and the ether group at O7 (Wang *et al.*, 1987; Odefey *et al.*, 1992; Mazzini *et al.* 1998). Adriamycin and daunomycin morpholino analogues have been shown to intercalate between the terminal 5′CG sequences of 5′d(CGATCG)$_2$ and 5′d(CGTACG)$_2$ (Odefey *et al.*, 1992; Mazzini *et al.*, 1998). These oligomers are particularly short and may not truly represent the molecular movement of β-DNA adducts of nuclear DNA. Terminal base pairs are not the site of choice to study the

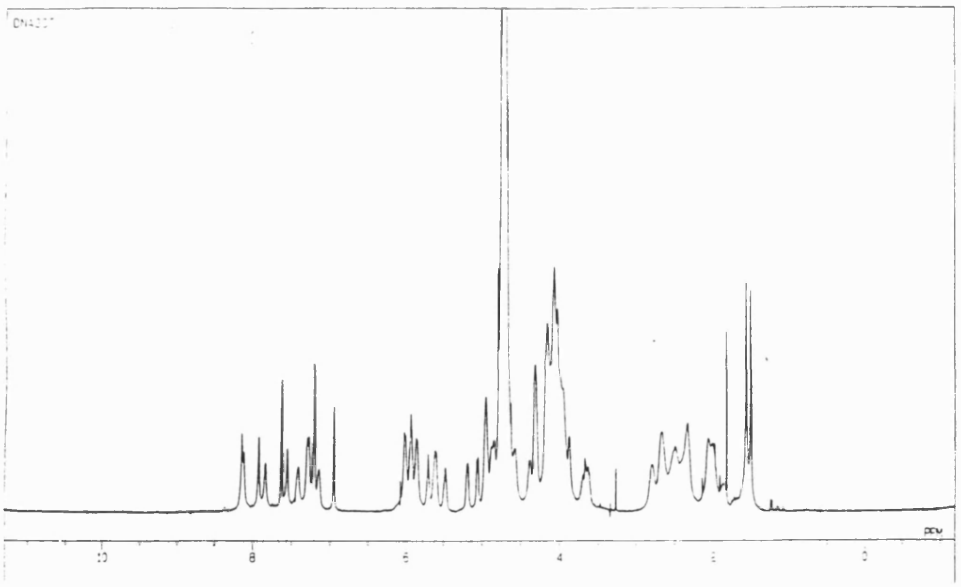


Figure 86 The 1D ^{1}H NMR spectrum of 5′d(GCTAGCTAIC)$_2$ DNA duplex.

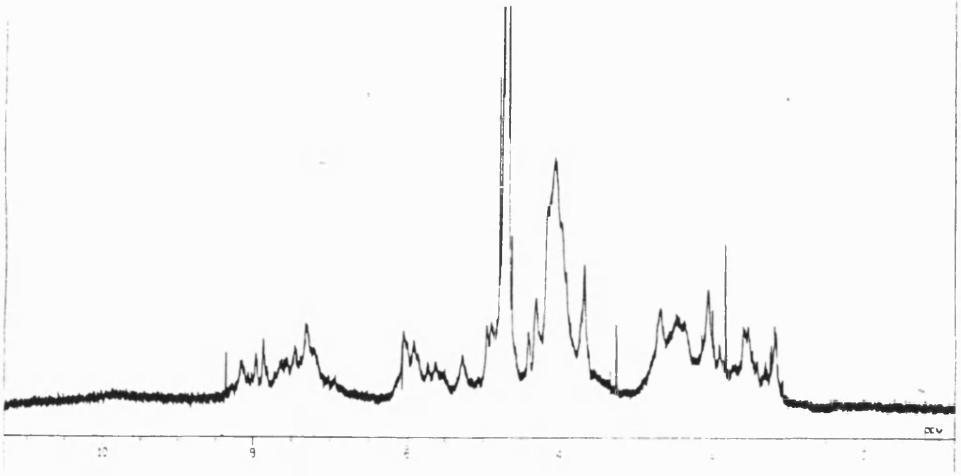


Figure 87 The 1D ^{1}H NMR spectrum of the CMA-5′d(GCTAGCTAIC)$_2$ reaction mixture, prior to purification.

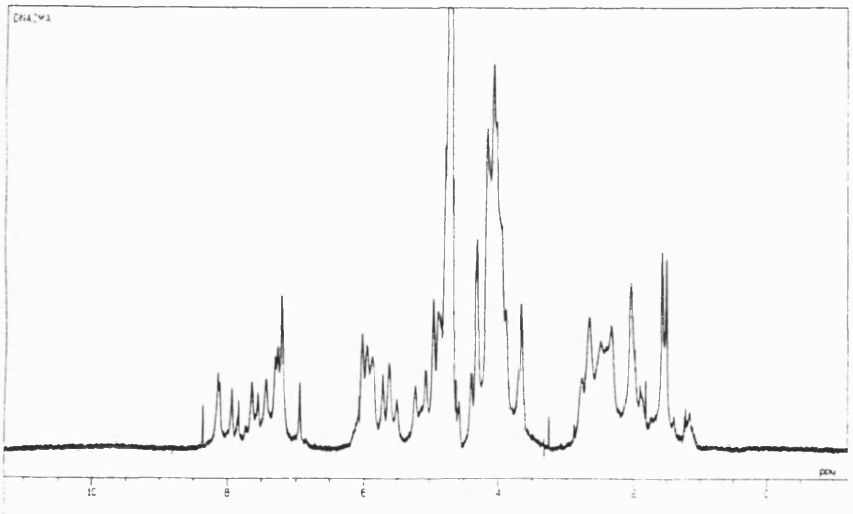


Figure 88 The 1D ^{1}H NMR spectrum post-purification of the CMA-5′d(GCTAGCTAIC)$_2$ reaction mixture.

effect of drug binding, because they are often subject to end-fray (base pair opening) and show stronger NOE connectivities than the bases in the centre of oligomers. This could confuse findings from NMR and molecular modelling studies. However, it would be interesting to see whether these analogues and, indeed, CMA would bind in a sequence-selective manner to the 5′CGTA or 5′CGAT consensus sequences when they are incorporated into the centre of a DNA duplex, or if they intercalate between 5′CG, projecting the morpholino 'tail' indiscriminately in both the 5′- and 3′- directions. This would then explain the use of the extremely short DNA hexamers which trap the ligand within the DNA minor groove in one orientation only. A species where the daunosamine and morpholino group project outside the helix would be energetically unfavourable. If CMA was unable to bind to 5′ATCGTA or 5′TACGAT for example, to form a single species owing to a lack of sequence specificity, this could also explain the multiple species obtained with the 5′GG containing duplex.

As with the design of the 5′d(GTTCCATGGAAC)$_2$ sequence, the make-up of the 5′d(GCTAGCTAIC)$_2$ was also flawed. There was no guanine within the second flanking base pair, either in the 5′ or 3′ direction, adjacent to the GC base pairs of the intercalation site. This would have presented an opportunity for covalent bond formation between the cyanomorpholino moiety of CMA and the nucleophilic N2 of guanine. An additional two bases added to the consensus sequence described by Chaires *et al.*, (1987) could possibly enhance the sequence selectivity of this reaction. After consideration of previous research and these data, the evidence seems to suggest that the structures in Figure 89,

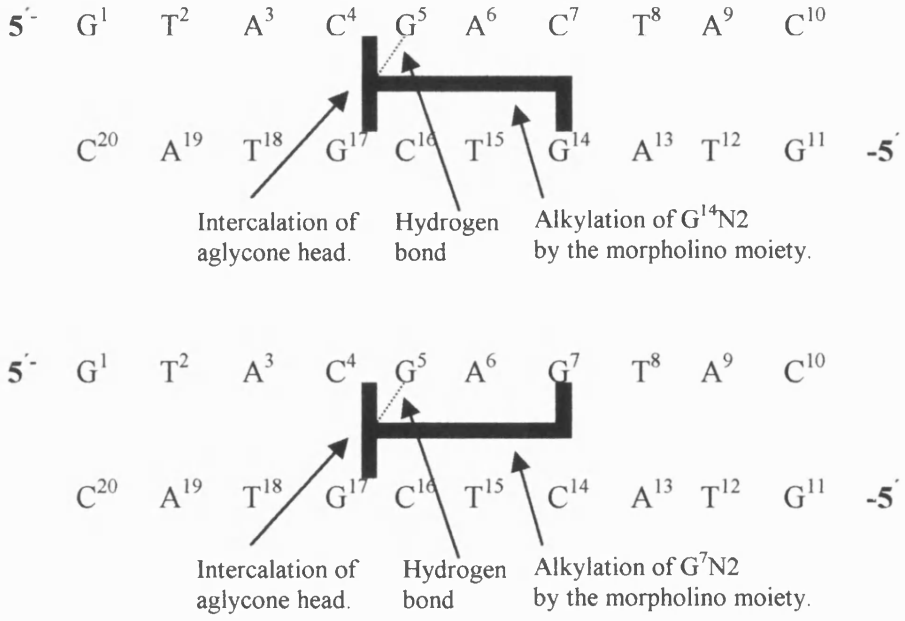


Figure 89 Possible DNA sequences producing CMA interstrand and intrastrand cross-links.

are likely candidates for interstrand and intrastrand cross-linkers respectively. Both sequences would hopefully lead to CMA-DNA monoadducts, if CMA does have a preferred binding sequence-orientation. The inability of such a potent minor groove binding agent as CMA to distinguish groove direction would be very unusual. This is because all other minor groove binding ligands (as discussed in Chapter 1) are extremely specific in their consensus sequence requirements, for enabling precise binding within the minor groove.

5.2 Intrastrand cross-linking in DNA by DSB-120

DSB-120 is a synthetic analogue of the naturally occurring pyrrolo[2,1-*c*][1,4]benzodiazepines (PBDs). This compound was allowed to react with the non-self-complementary DNA duplex 5′d(GTGATGAG)·5′d(CTCATCAC), in the hope of achieving an intrastrand cross-linked adduct.

5.2.1 The 1D ^{1}H NMR spectrum of 5′d(GTGATGAG)·5′d(CTCATCAC)

Owing to the non-palindromic nature of this sequence, each strand of this duplex had to be synthesized separately. A one-dimensional proton NMR spectrum was collected on each strand, to check the purity of the sample and that the correct oligonucleotide had been synthesized.

Figure 90 shows the spectrum of the pure 5′d($G^1T^2G^3A^4T^5G^6A^7G^8$) octamer. Two methyl resonances were noted at δ 1·55 and δ 1·6ppm, which correlated with the two thymines (T^3 and T^5) present in this sequence. Examination of the aromatic region between δ 7 and δ 8·5ppm showed four clear singlets at δ 7·1, δ 7·25, δ 8·0 and δ 8·2 ppm. The two peaks furthest upfield are likely to correspond to thymine (T^3 and T^5), owing to their weaker ring current effects yielding greater shielding than seen with purine aromatic protons. A^4 and A^7H8 are more deshielded so must correspond to the peaks furthest downfield at δ 8·0 and δ 8·2ppm. The peaks in the centre of this region are quite overlaid but must incorporate G^1, G^3, G^6 and G^8H8 protons, plus two aromatic AH2 protons.

Figure 91 displays the ^{1}H NMR spectrum of the oligomer 5′d($C^9T^{10}C^{11}A^{12}T^{13}C^{14}A^{15}C^{16}$). Inspection of the methyl region shows two large singlets at δ 1·6 and δ 1·8ppm, consistent with the presence of two methyls; T^{10} and T^{13}. Analysis of the aromatic region shows a large number of well-resolved peaks. Protons with shifts of approximately δ 8·35, δ 8·3, δ 8·1 and δ 8·0 ppm are most likely to correspond to adenine H8 and H2 protons of A^{12} and A^{15}. The peak at δ 7·3ppm is probably due to a thymine H6 (T^{10} or

T^{13}). The remaining peaks are more overlaid but it is possible to discern four sets of doublets at δ 7·45, δ 7·5, δ 7·7 and δ 7·75ppm which will tally with cytosine C^9, C^{12}, C^{14} or C^{16}H6s. The last peak at δ 7·6ppm should correspond to the remaining thymine (T^{10} or T^{13}) H6.

These samples were then mixed together and annealed to form a β-DNA duplex. Hydroxylapatite chromatography was used to separate the double-stranded DNA from any excess single-stranded DNA. The NMR spectrum of the 5′d(GTGATGAG)·5′d(CTCATCAC) duplex is given in Figure 92. Small changes in chemical shift were noted for many resonances on duplex formation. A total of four methyl signals at δ 1·3 (x2 overlaid), δ 1·4 and δ 1·6ppm were noted, relating to the methyl protons of nucleotides T^3, T^5, T^9 and T^{13}. The aromatic region is more crowded now that it harbours 4 x TH6, 4 x AH8, 4 x AH2, 4 x GH8 and 4 x CH6 (doublets) proton resonances. This gives a total of 20 predicted peaks, 4 of which are split into doublets, which equals the peak count in this region of the spectrum.

5.2.2 DSB-120 may form an intrastrand cross-link in the DNA sequence 5′d(GTGATGAG)·5′d(CACTACTC

DSB-120 is a synthetic antitumour agent. Its structure consists of two C8-linked pyrrolo[2,1-*c*][1,4]benzodiazepine subunits, each of which are capable of monofunctional alkylation of DNA *via* C11a and the N2 of guanine (Hurley *et al.*, 1977). The formation of a dimer results in bifunctional alkylation of the DNA, yielding interstrand cross-links. Jenkins *et al.*, (1994) and Mountzouris *et al.*, (1994) produced a refined molecular model, from NMR data, of the symmetric cross-link between G^4 (one on each strand) of the self-complementary DNA duplex 5′d(CICG*ATCICG)$_2$. DSB-120 was shown to span a total of six base pairs at its preferred consensus sequence 5′(Pu/Py)G*A TC(Py/Pu), where * and _(on the opposite strand) represent sites of covalent modification.

The 5′d(GTGATGAG)·5′d(CACTACTC) DNA duplex contained only the first three bases of the 5′d(GATC) consensus sequence (Jenkins *et al.*, 1994). This was because a guanine was required at position four instead of a cytosine, to provide an opportunity for the formation of an interstrand DNA-DNA cross-link *via* DSB-120.

Figure 93 displays the one-dimensional proton NMR spectrum of 5′d(GTGATGAG)·5′d(CACTACTC) after it had been allowed to react with DSB-120 and purified using C_{18} Sep Pak cartridges. Although the sample is not very concentrated owing to some loss during purification, it is evident from the increased number of peaks in this spectrum that some reaction has taken place. Examination of the methyl region (Appendix XX Fig. 178 gives an expansion) shows there are approximately six peaks at δ 1·6, δ 1·55, δ 1·4, δ 1·35, δ 1·3

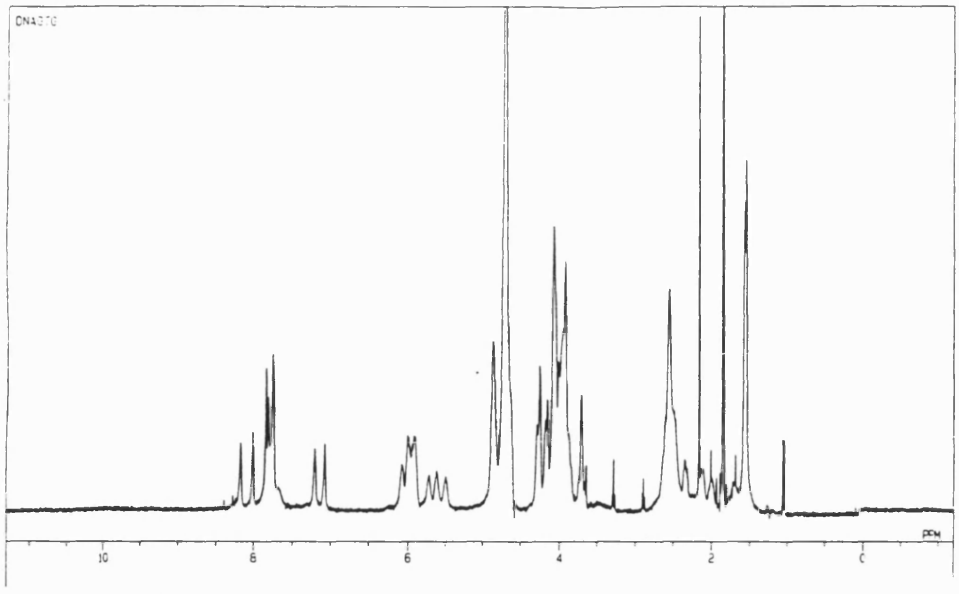


Figure 90 The 1D ^{1}H NMR spectrum of the single strand 5′d(GTGATGAG) DNA oligomer.

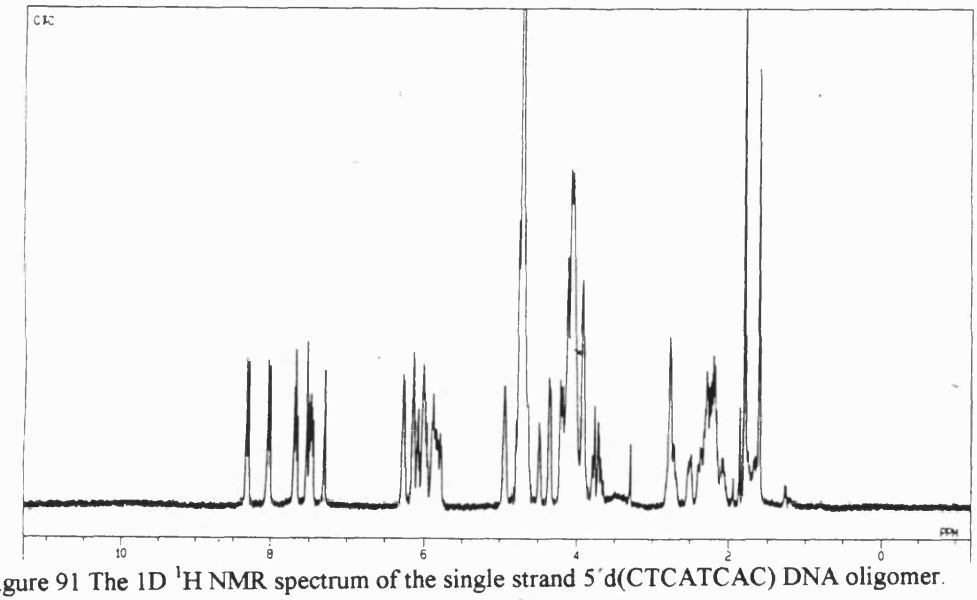


Figure 91 The 1D ^{1}H NMR spectrum of the single strand 5′d(CTCATCAC) DNA oligomer.

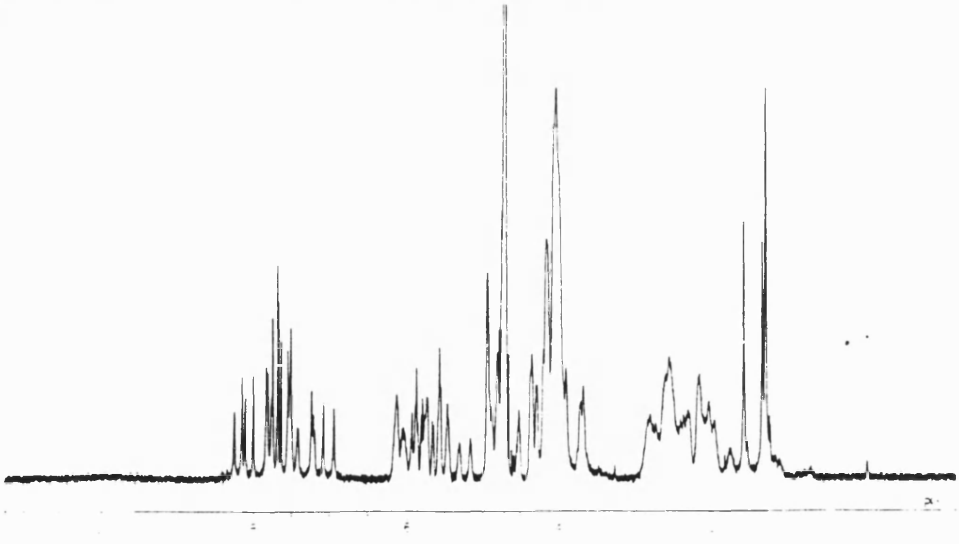


Figure 92 The 1D ^{1}H NMR spectrum of 5′d(GTGATGAG)·5′d(CTCATCAC) DNA duplex.

and δ 1·15ppm. Three peaks seem to represent unmodified DNA, duplex signalling that the reaction with DSB-120 is not at completion (perhaps only around 40% from the ratio of peak heights). Further exposure of the DNA to DSB-120 did not change the topology of the spectrum, indicating that equilibrium had been attained. Two new methyl peaks at δ 1·35 and δ 1·15ppm correspond to the DSB-120 modified DNA. The remaining two methyls of the DNA which has reacted with DSB-120 seem to be overlaid with duplex resonances, perhaps indicating little change in chemical environment for these TCH_3.

Analysis of the aromatic region not only shows a large increase in the number of peaks but, more interestingly, a broad hump has appeared at about δ 8·7ppm. Mountzouris *et al.*, (1994) noted a similar peak in 90% D_2O at 9·1ppm in the DSB-120-5′d(CICGATCICG)$_2$ adduct. This was attributed to the NH proton of guanine N2 which is covalently bound to DSB-120 at C11, *via* an aminal linkage. It was thought to be a result of DSB-120-induced conformational stress within the minor groove, making it more accessible to the solvent. The fact that this peak is present in Figure 93 supports covalent modification of 5′d(GTGATGAG)·5′d(CACTACTC) by DSB-120. It is not possible to tell whether this ligand has only monoalkylated the DNA, probably at G^3 which is incorporated in its preferred consensus sequence 5′(Pu/Py)G^3A^4, without additional two-dimensional ^{1}H NMR NOESY, DQF-COSY and deuterium exchange data. It is also conceivable that an intrastrand cross-link may have been formed between G^3 and G^6, as hoped (Fig. 92).

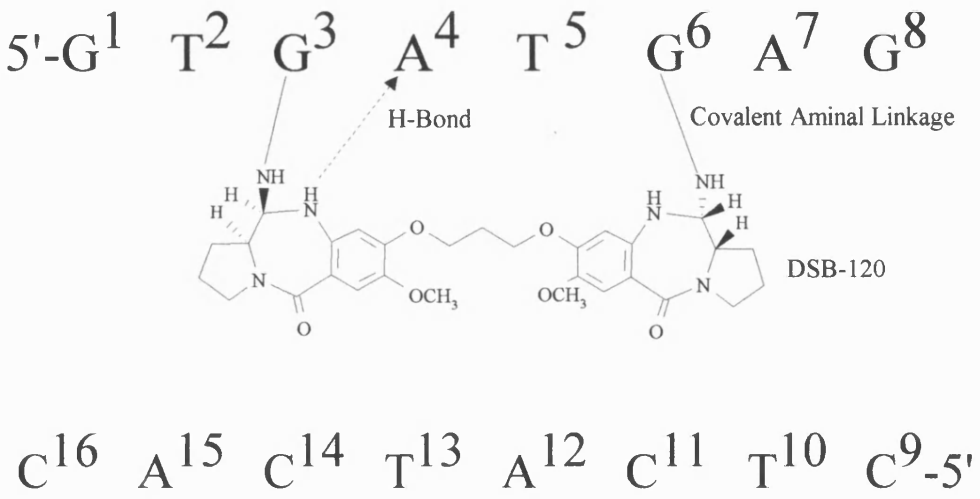


Figure 92 The proposed structure of the intrastrand DNA-DNA cross-link formed by DSB-120.

However, the fact that this reaction did not proceed to completion indicates that there is either a problem with the target sequence of the oligomer, or that intrastrand cross-linking has an incurred energy penalty, due to the deviation of the propanediyldioxy linkage from its preferred

trans conformation, as seen in the interstrand cross-link. No further work was possible to investigate the exact nature of the observed interaction between DSB-120 and this DNA sequence, owing to the degradation of the PBD dimer.

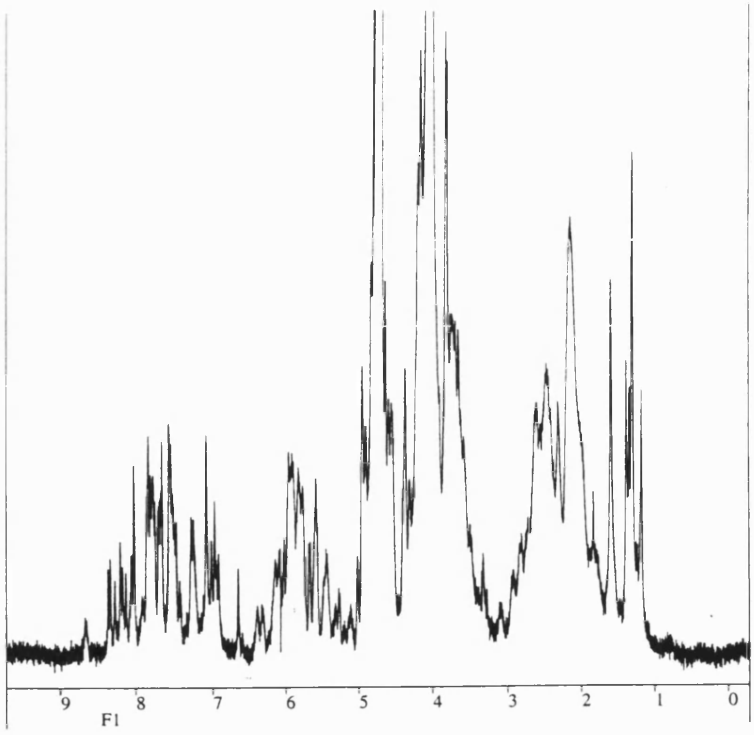


Figure 93 The 1D ^{1}H NMR spectrum of 5′d(GTGATGAG)·5′d(CACTACTC) after reaction with DSB-120.

5.3 The reaction of sibiromycin with DNA

Sibiromycin is a pyrrolo[2,1-*c*][1,4]benzodiazepine (Gause *et al.* 1969). This drug was allowed to react with the palindromic DNA sequences 5′d(GAGATCTC)$_2$ and 5′d(GCAGAATTCTIC)$_2$ to try to produce a DNA adduct *via* alkylation of G-N2 (Hurley *et al.* 1977).

5.3.1 1D ^{1}H NMR spectra of 5′d(GAGATCTC)$_2$ and 5′d(GCAGAATTCTIC)$_2$

This NMR spectrum (Fig. 94) of 5′d(G^{1}A^{2}G^{3}A^{4}T^{5}C^{6}T^{7}C^8)$_2$ was analysed as previously described. The methyl region contained only one large singlet which integrated for six protons, corresponding to thymines T^5 and T^7 which were overlaid at δ 1·6ppm. The two upfield resonances in the aromatic region (δ 7·05 and δ 7·2ppm) could be attributed to T^5 and T^{7}H6. In contrast, the two most deshielded protons were likely to be adenine A^2 and A^{4}H8s. The remaining peaks in this region were overlaid, owing to their integrals being twice the height of

the adenine and thymine singlets. This accounted for the remaining guanine and cytosine protons.

The 5′d($G^1C^2A^3G^4A^5A^6T^7T^8C^9T^{10}I^{11}C^{12}$)$_2$ sequence produced the NMR spectrum shown in Figure 97. Three methyl resonances were noted at δ 1·15, δ 1·4 and δ 1·55ppm, which corresponded to thymines T^7, T^8 and T^{10}. There appears to be the correct number of aromatic H6 and H8 protons equal to: 2 x GH8, 3 x CH6, 3 x AH8, 1 x IH8 and 3 x TH6. Inosine and adenine H2 protons appear to be overlaid with the H6/H8 peaks.

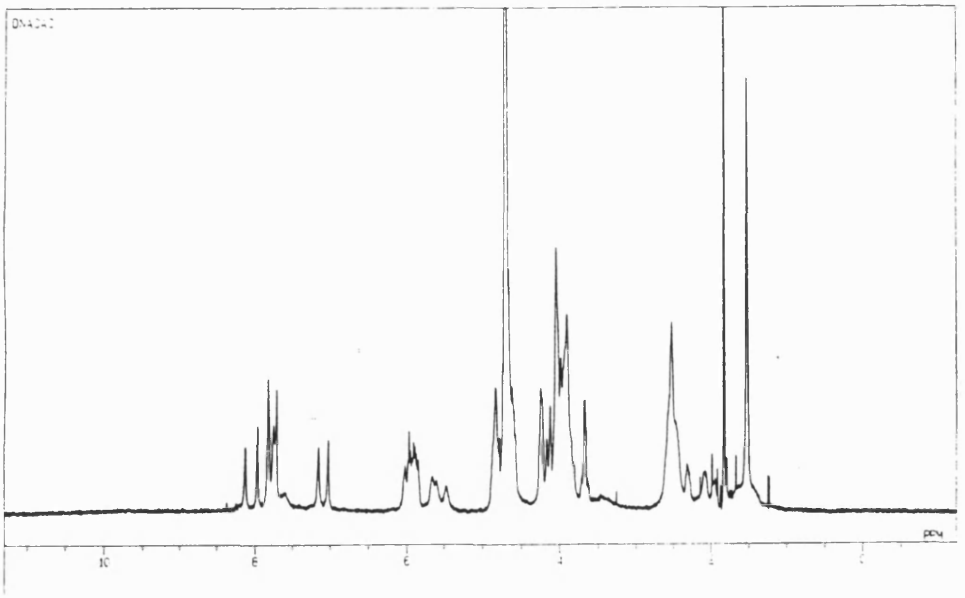


Figure 94 The 1D ^{1}H NMR spectrum of 5′d(GAGATCTC)$_2$.

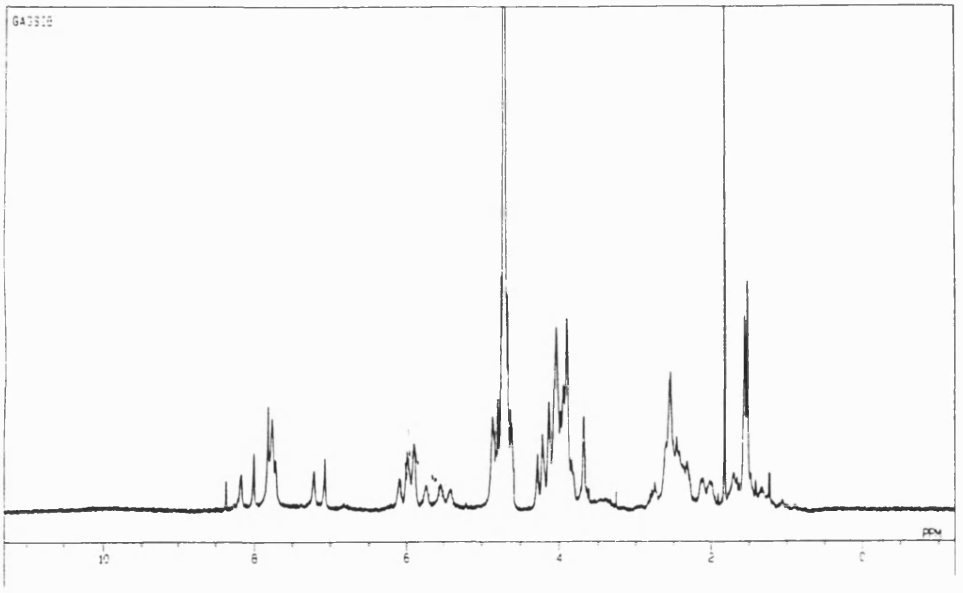


Figure 95 The 1D ^{1}H NMR spectrum following exposure of 5′d(GAGATCTC) to sibiromycin.

5.3.2 The reaction of sibiromycin with 5′d(GAGATCTC)$_2$ and 5′d(GCAGAATTCTIC)$_2$

The methyl region of the NMR spectra yielded from each oligonucleotide sequence would be expected to symbolize whether sibiromycin had indeed formed a drug-DNA adduct on mixing. Purification using C_{18} Sep Pak cartridges would have removed any unreacted drug from the sample. Additional methyl resonances observed in these spectra, therefore, would symbolize sibiromycin covalently bound to DNA. There are five methyl groups in the structure of sibiromycin (Fig. 96), whose expected chemical shifts would be approximately: $\underline{CH_3}$-C=C δ=1·6, $\underline{CH_3}$-OAr=δ=3·8, $\underline{CH_3}$-C-O δ=1·3, $\underline{CH_3}$-R δ=0·9 and $\underline{CH_3}$-N δ=2·3ppm (Williams and Flemming, 1987).

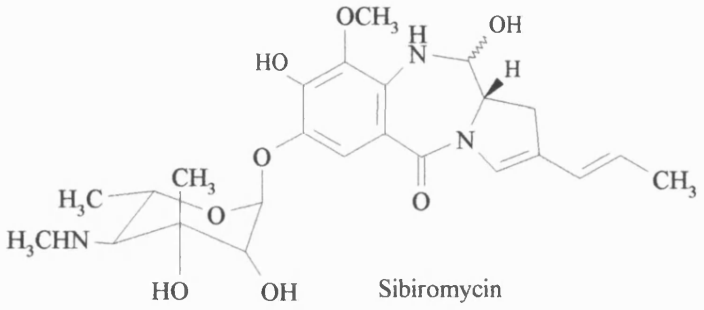


Figure 96 The molecular structure of sibiromycin.

Figure 95 shows that no modification of the DNA self-complementary sequence 5′d(GAGATCTC)$_2$ has occurred, even after 2 x 48h exposures to sibiromycin. This ^{1}H NMR spectrum is identical to that of the duplex DNA in Figure 94. Sibiromycin does not bind to the synthetic oligomer 5′d(GAGATCTC)$_2$.

Inspection of Figure 98, however, shows there is a marked increase in the complexity of this ^{1}H NMR spectrum, symbolizing a possible reaction between 5′d($G^1C^2A^3G^4A^5A^6T^7T^8C^9T^{10}I^{11}C^{12}$)$_2$ and sibiromycin. Comparison of this spectrum with the one corresponding to the duplex (Figure 97), demonstrated that the reaction of this sequence with sibiromycin had not achieved completion yielding adduct only. The methyl resonances corresponding to T^7, T^8 and T^{10} were in analogous positions in both spectra, indicating duplex DNA persisted after two exposures to sibiromycin. Some new methyl peaks did appear in Figure 98, which corresponded to both sibiromycin methyls, as well as those of T^7, T^8 and T^{10}. The chemical environments of these thymine methyls are likely to have changed on modification of this sequence by sibiromycin, leading to upfield or downfield shifting of these peaks to new positions. These new methyl peaks seem to be located at approximately δ 1·15(2 x CH_3 overlaid?), δ 1·35, δ 1·6, δ 1·65 and δ 1·75ppm. They are all only about two thirds of the

height of the remaining DNA duplex methyl peaks, some existing as shoulders only on these peaks. This suggests that sibiromycin has at best, only modified 40% of the DNA in this sample *i.e.* a ratio of 3:2 unreacted:reacted DNA. No further analysis of the nature of the sibiromycin-modified DNA was undertaken owing to the small size of the sample, which would have made 2D ^{1}H NOESY assignments extremely difficult.

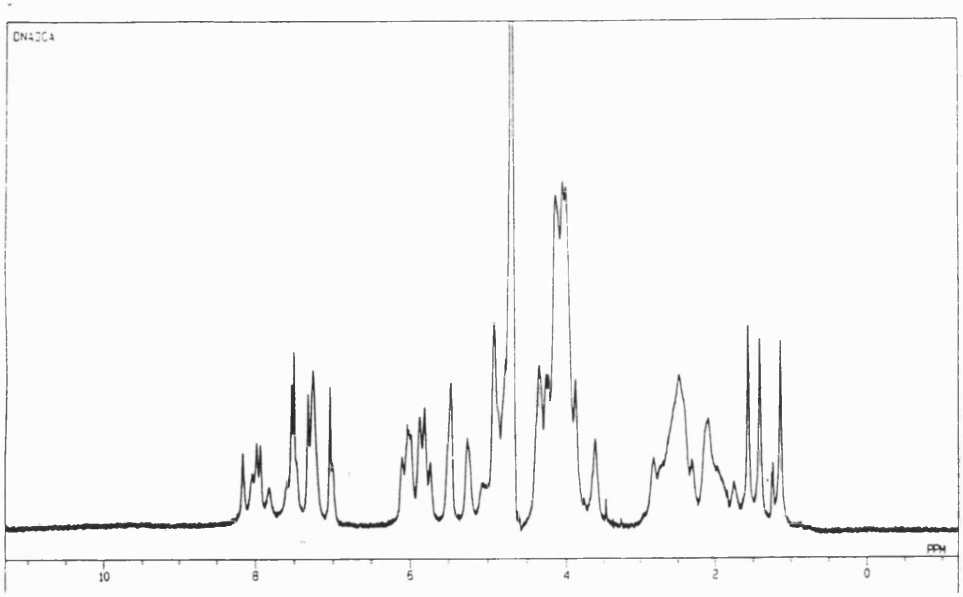


Figure 97 The 1D ^{1}H NMR spectrum of 5′d(GCAGAATTCTIC)$_2$.

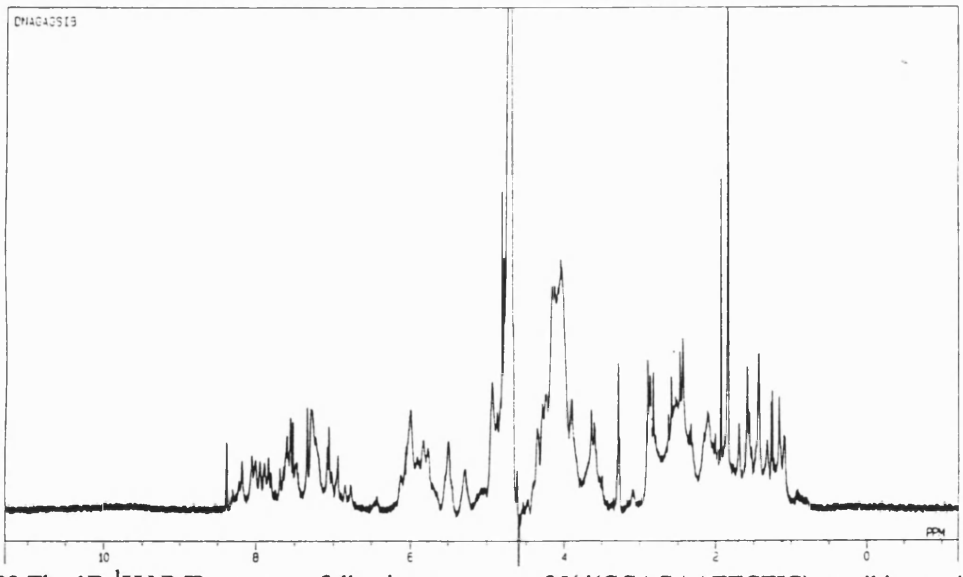


Figure 98 The 1D ^{1}H NMR spectrum following exposure of 5′d(GCAGAATTCTIC)$_2$ to sibiromycin

Hurley *et al.*, (1980) constructed CPK models to demonstrate how the concave surface (C9, N10, C11, C11a and C1) of PBDs such as sibiromycin bound isohelically within the minor groove. Mostad *et al.*, (1978) described the stabilizing influence the sibirosamine sugar would have from within the minor groove *via* its interaction with the sugar phosphate backone. Sibiromycin can compete for the binding sites of anthramycin and tomaymycin (Hurley *et al.*,

1977). The most preferred binding sequence in DNA for the PBDs is 5′PuGPu (Hertzberg *et al.*, 1986). Pierce *et al.*, (1993) defined these purines as adenines in the case of anthramycin and tomaymycin binding, yielding the consensus sequence 5′AGA. Considering sibiromycin can compete for both of these binding sites, it would be logical that 5′AGA would also be a consensus sequence for sibiromycin. Puvvada *et al.*, (1997) established that in fact sibiromycin was a stronger inhibitor of transcription by binding to plasmid DNA, than tomaymycin, anthramycin or neothramycin. Blockage of transcription took place at the 5′AGA sequence by sibiromycin.

Considering the preferred consensus sequence for sibiromycin binding has proved to be 5′AGA, the 5′d(GAGATCTC)$_2$ and 5′d(GCAGAATTCTIC)$_2$ DNA duplexes used in this study, would have been expected to produce sibiromycin-DNA adducts. The latter sequence did undergo some low level modification, perhaps due to the higher adenine content within the flanking sequences (Puvvada *et al.*, 1997). It is probable that the consensus sequence for sibiromycin is more complex than that described by a 5′AGA trimer. The flanking sequence could induce some conformational effect on DNA which makes the binding of sibiromycin more favourable. It would be interesting to see if synthetic DNA oligomers consisting of base combinations such as 5′AAAGACA or 5′CAAGAAC, which were pinpointed as particularly reactive sequences in the transcription footprinting assays, were more successful in forming synthetic sibiromycin-DNA adducts.

It is also possible that some environmental factor may have inhibited sibiromycin binding to these synthetic oligomers. Gause *et al.*, (1975) found that sodium dodecylsulphate completely blocks the interaction of sibiromycin with DNA, but is unable to induce dissociation of the ligand from a preformed complex. Dimethylformamide has also been demonstrated to be unable to remove sibiromycin from a preformed drug-DNA adduct. However, dimethylformamide was added to the sibiromycin/NMR buffer/D_2O/DNA reaction mixture to increase the solubility and hence, availability of sibiromycin for reaction. Dimethylformamide could be responsible for blocking sibiromycin-DNA adduct formation. Gause and Dudnik, (1972) described the inhibition of the sibiromycin reaction with native DNA in the presence of Mg^{2+} ions. Some ion contamination in these samples may have produced the same effect. Use of a chelating agent such as EDTA to sequester ions, could solve a problem such as this.

It is not clear why sibiromycin did not react with these synthetic nucleotides containing the preferred 5′AGA consensus sequence. This is particularly perplexing when sibiromycin is known to be a potent alkylator of DNA.

CHAPTER 6: CONCLUSION

The covalent modification of two DNA sequences 5′-d(CGTAAGCGCTTA*CG)$_2$ and 5′-d(CGAAAAA*CGG)·5′-d(CCGTTTTTCG) by adozelesin has been investigated by high-field NMR. A refined molecular model was generated of each adozelesin-DNA adduct structure using distance-range restraints. The molecular models of these DNA adducts which demonstrate the binding of the pharmacologically active, monofunctional cyclopropapyrroloindole adozelesin are the first of their kind.

Despite the differences in these DNA sequences, there were some fundamental similarities between the two adduct structures. Adozelesin covalently modified the N3 of adenine from within the minor groove, with the benzofuran 'tail' extending in the 5′-direction. In each case, the concave hydrophobic surface of the drug was directed deep into the minor groove, whilst the convex hydrophilic spine was directed out of the groove in contact with the surrounding solvent. The sequence-specificity of adozelesin has been previously noted in sequencing studies (Weiland and Dooley, 1991), who described p(A) and 5′-(T/A)(TA)T-A*(C/G)G consensus sequences. The 5′-TTA*CG and 5′-AAAAA* binding sites were incorporated into the DNA duplexes used in this study and, in both cases, adduct formation was sequence specific, yielding only one major adduct species.

Perhaps one of the most interesting observations noted within the 5′-d(CGTAAG^{6}CGCTTA*CG)$_2$-adozelesin adduct is the significant distortion noted around the central guanine (G^6). This guanine is on the unmodified strand and is adjacent to the ureylene linker between the benzofuran and indole sub-units of adozelesin. Previous studies with bizelesin (Thompson *et al.*, 1995) have indicated the important role of this linker in determining the sequence specificity of this dimeric compound. A hydrogen-bond from the exocyclic amine of the central guanine in the consensus sequence to the carbonyl of the linker distorted the duplex in such a manner that seven base-pairs (5′-TTAGTTA) were spanned to form the cross-link instead of six. This NMR and dynamics study indicates that a similar situation also exists within the monomeric CPI analogues, in that this linking function between the sub-units may have a significant effect in allowing the binding of these ligands to DNA sequences which would not normally be favourable for binding. Whilst Weiland and Dooley (1991) restricted their consensus sequence to AT base pairs, Yoon and Lee (1998) have identified the incorporation of a GC base-pair at the first and second flanking base-pair positions. GC-rich minor grooves are generally wider, with the steric problem of the guanine-NH_2 projecting into the minor groove and colliding with any ligand situated there. The hydrogen-bond between the amide of the linker and the carbonyl of thymine possibly tethers the adozelesin to the modified strand, stabilizing the structure so that the benzofuran 'tail' does not 'flail' around within the

minor groove. The second, weak (or maybe even transient) bifurcating hydrogen-bond to the guanine-NH_2 may encourage the observed distortion around this base.

However, a second strong explanation for the distortion around G^6 is that it is a result of a distortional wave propagated from 6 bp upstream where a second adozelesin molecule is alkylating this duplex at A^{12} on the same strand. Similar proposals of this nature have been cited in the literature previously. Seaman *et al.*, (1996) established that CPI-I bisalkylation of 5′-TAA^2AAA^1 proceeded in a 'head to tail' fashion as a result of this phenomenon. Lin *et al.*, (see review by Hurley and Draves, 1993) also observed a distortion around a cytosine ($G^7 \equiv C^{18}$) adjacent to (+)CC-1065 on the non-covalently modified strand of the sequence 5′-$G^7T^8T^9A^{10*}$. This cytosine is in exactly the same position as G^6 of this study. A similar conformational change has also been noted in homeodomain binding specificity studies (Ades and Sauer, 1994). The results of this study verify that cyclopropapyrroloindoles give rise to a distortion when binding to the minor groove. Without further investigation, one can only speculate whether it is upstream of the modification site or adjacent to the ureylene linker as a result of a bifurcating bond. The CPI-I ligand does not contain a ureylene linker owing to the lack of a third sub-unit. Therefore, the distortion noted within the CPI-I bisadduct cannot be attributed to the ureylene linker. This suggests the former explanation is more favourable *i.e.* it is the cyclopropapyrroloindole 'headunit' common to all of these structures which is responsible for the distortion. This could be examined by binding a CPI 'headunit' only to DNA duplex to form a monoadduct, using a duplex sequence such as 5′-d(CGCGCGTTA*CGCGCG)$_2$. This duplex contains a single alkylation site (due to its self-complementary nature) and eight flanking base-pairs in the 5′-direction to detect any distortion. No evidence was available from the 'A' tract adduct to determine whether distortion was present, because the 6^{th} flanking base-pair was a terminal. Extending the terminal ends by an extra set of two base-pairs would provide an answer to this question.

Previous studies with (+)CC-1065 have also shown that the ethano bridges of the B and C sub-units in the natural product are responsible for the observed delayed lethality and DNA over-winding (Sun and Hurley, 1992b; Warpehoski *et al.*, 1988). Interestingly, in (+)CC-1065, the NH of the amide linker is part of a ring system incorporating these ethano bridges and the amine is hence unable to interact with the DNA, as observed in this self-complementary adozelesin-DNA adduct. Obviously, the origins of over-winding must originate from ligand-DNA interactions around the amide linker between the sub-units. A future direction of this project could be to synthesize a closely related (+)CC-1065 and adozelesin duplex-adduct using the same DNA sequence, to try to establish the exact structural origins of the DNA over-winding in (+)CC-1065 adducts. The observation of significant hydrogen-bonding within the adozelesin-self complementary DNA adduct would appear to indicate a significant difference in

the binding mechanism for these two closely structurally related compounds. It is likely that this difference holds the key to the inability of adozelesin to over-wind DNA and cause the delayed lethality observed with (+)CC-1065.

The fact that the adozelesin-5′-d(CGTAAGCGCTTA*CG)$_2$ adduct results in a symmetrical bisadduct with adozelesin orientated 'tail to tail' yields the possibility of synthesizing an adozelesin dimer. The bisalkylation of 5′-TAA^{2}AAA^{1} by CPI-I (half of bizelesin dimer) proceeds unsymmetrically in a 'head to tail' orientation of the ligands. This indicates a conformational change in the second binding site (6 bp upstream) after the first CPI-I binds, which leads to the second CPI-I binding more favourably in the reverse orientation. A distortion 6 bp to the 5′-side of the first binding site to which bizelesin alkylates is likely to make the second alkylation less favourable, which is reflected in the rate of cross-link formation (Woynarowski *et al.*, 1995). The adozelesin bisadduct does not encounter this problem; hence, it would follow that a synthesized adozelesin dimer, wherein the adozelesin molecules are linked 'tail to tail' *via* a linker similar to that of bizelesin, would produce an extremely potent cross-linker with increased sequence selectivity, owing to the drug being able to span a longer consensus sequence of 10 base-pairs. This is approaching the 15 base-pairs required for recognition of a single site on the human genome, for complete sequence selectivity.

It is clear from inspection of the DNA 'A' tract-adozelesin adduct that binding of adozelesin within this sequence yields a complementary fit better than that observed in the self-complementary adduct. Within the 'A' tract, adozelesin lies 'edge-on' and centrally positioned along the length of this deep, narrow minor groove. The facts that the sub-units of adozelesin do not show alternate close associations with the modified and unmodified strands or curving of the drug demonstrates why adozelesin is able to span six rather than five base-pairs. The NMR data and the resulting refined molecular model both indicate that 'A' tract character has been maintained after formation of the adduct. This is the first monofunctional cyclopropapyrrolo-indole-'A' tract DNA adduct to be characterized.

The maintenance of propellor twisted 'A' tract character (Nelson *et al.*, 1987) was detected after adduct formation by high-field NMR. The insertion of an 'A' tract sequence (the length of a half-helical turn) between two general base sequences with no base-pair roll yields a macroscopically bent DNA structure (Nelson *et al.*, 1987; Goodsell *et al.*, 1994). Although β-helical structure is conserved on adduct formation, it is obvious from the model that the DNA is macroscopically bent, with the flanking base pairs to the 3′-side of the alkylation site showing the most deviation from the helix axis. A slight compression of the major groove is also observed. Adozelesin clearly favours the unique conformation of this DNA sequence. The bending of the helix, beginning at the end of the 'A' tract, facilitates the fit of the CPI headunit,

with the projecting methyl allowed more space to follow the floor of the minor groove as the helix bends. This allows adozelesin to bind perfectly 'edge-on' and in a central position. The narrowness of the minor groove promotes many non-covalent interactions (van der Waals forces) from the close complementarity of fit of adozelesin, which enhances the bias of the sequence selectivity of the drug towards 'A' tract sequences. This agrees with the finding of Weiland and Dooley (1991), where p(A) was the most preferred consensus sequence, as indeed it is of many minor groove binding agents for the same reasons. Unlike previous studies where the binding of bizelesin was able to trap out straight DNA after the formation of an interstrand cross-link across the 'A' tract impeded propellor twisting (Thompson and Hurley, 1995), monoalkylation does not affect 'A' tract character which is advantageous to binding.

Only multiple species of the cyanomorpholinoadriamycin-DNA adduct were obtained with the DNA sequences used in this study. Hence, no refined molecular model was obtainable. However, an adduct has been reported in the literature (Mazzini *et al.*, 1998) where CMA is intercalated between the terminal base-pairs of the 5′-d(CGATCG)$_2$. This was not thought to give a true representation of the adduct structure in an NMR study, owing to the effects of fraying of the terminal bases. It would be interesting to see whether CMA would bind in a sequence selective manner to the 5′-CGTA or 5′-CGAT consensus sequences when they are incorporated into the centre of a DNA duplex. The use of a terminal strand intercalation site may be a tactic to avoid CMA binding to the minor groove in either orientation. A sequence-selective minor groove binder which did not specify a particular binding orientation, *i.e.* the morpholino 'tail' directed to 5′- or 3′-, would be very unusual. The sequences 5′-d(GTACGACTAC) and 5′-d(GTACGAGTAC) could be a good starting point for producing interstrand and intrastrand cross-linking monoadducts, respectively.

The pyrrolo[2-1-*c*][1,4]benzodiazepines DSB-120 and sibiromycin were also investigated. An NMR study on the DSB-120-5′-d(GTG*ATG*AA) adduct suggested that an intrastrand cross-link may have formed in this sequence. This would be particularly interesting because DSB-120 is normally associated only with interstrand cross-links. Sibiromycin, however, could be persuaded to react with the 5′-d(GCAGAATTCTIC)$_2$ duplex only partly, while it did not react at all with the 5′-d(GAGATCTC)$_2$ duplex. This was quite unexpected, considering that both sequences contain the 5′-AGA site which is known to be the consensus sequence for sibiromycin (Hurley *et al.*, 1977; Hertzberg *et al.*, 1986). This suggests that the consensus sequence for sibiromycin is more complex than the 5′-AGA trimer. It is possible that the flanking sequence, *e.g.* like the 'A' tract in the adozelesin adduct, induces some conformational change of DNA which makes the binding of sibiromycin more favourable.

It is clear from the varying degrees of success in the production of drug-DNA adducts investigated in this project that the requirements of the make-up of the DNA consensus sequence as defined by the drug are extremely stringent. These minor groove binding ligands will not be 'fooled' into binding to any base in DNA just because it can form a covalent bond to it. The complementarity of the fit between the ligand and the binding site and, more importantly, the abundance of non-covalent interactions to which it gives rise within the flanking sequence, dictates the sequence-specificity of these minor groove binding drugs.

The adozelesin-DNA self-complementary adduct in this study has provided a new and exciting starting point for rational drug design of an adozelesin dimer, a proposed antineoplastic agent with enhanced sequence-selectivity. Not only has the first CPI-'A' tract adduct structure been characterized where adozelesin is a perfect isohelical fit within the minor groove, but it has also been demonstrated how adozelesin achieves tolerance of unfavourable GC-bases within its consensus sequence. These adozelesin-DNA adducts have proved to be an excellent model system to investigate the molecular basis for DNA sequence recognition and binding. An understanding of sequence-specificity is the key first step in the design of new drugs which will react with their molecular target in a defined and predictable manner.

APPENDIX I

PERKIN ELMER 62 STEP PROTOCOL FOR OLIGONUCLEOTIDE SYNTHESIS

Step Number	Function Number	Function Name	Step Time
1	10	#18 to waste	6
2	9	#18 to Column	360
3	2	Reverse Flush	90
4	1	Block Flush	6
5	28	Phos Prep	5
6	61	Tet to Waste	3
7	19	B+Tet to Column	100
8	4	Wait	30
9	10	#18 To Waste	10
10	9	#18 To Column	50
11	2	Reverse Flush	80
12	1	Block Flush	4
13	16	Cap Prep	10
14	22	Cap to Column	60
15	2	Reverse Flush	5
16	22	Cap to Column	50
17	2	Reverse Flush	5
18	22	Cap to Column	50
19	4	Wait	120
20	10	#18 To Waste	10
21	1	Block Flush	6
22	81	#15 To Waste	10
23	13	#15 To Column	180
24	4	Wait	30
25	10	#18 To Waste	10
26	2	Reverse Flush	80
27	1	Block Flush	5
28	9	#18 To Column	100
29	2	Reverse Flush	80
30	9	#18 To Column	100
31	2	Reverse Flush	80

32	9	#18 To Column	100
33	2	Reverse Flush	80
34	9	#18 To Column	100
35	2	Reverse Flush	80
36	1	Block Flush	6
37	33	Cycle Entry	1
38	10	#18 To Waste	7
39	9	#18 To Column	50
40	2	Reverse Flush	80
41	1	Block Flush	6
42	6	Waste Port	1
43	5	Advance FC	1
44	82	#14 To Waste	5
45	14	#14 To Column	80
46	2	Reverse Flush	5
47	14	#14 To Column	50
48	2	Reverse Flush	5
49	14	#14 To Column	50
50	2	Reverse Flush	5
51	14	#14 To Column	50
52	2	Reverse Flush	5
53	14	#14 To Column	50
54	2	Reverse Flush	5
55	14	#14 To Column	50
56	1	Block Flush	6
57	10	#18 To Waste	5
58	9	#18 To Column	160
59	2	Reverse Flush	80
60	1	Block Flush	6
61	7	Waste Bottle	1
62	34	Cycle End	1

APPENDIX II

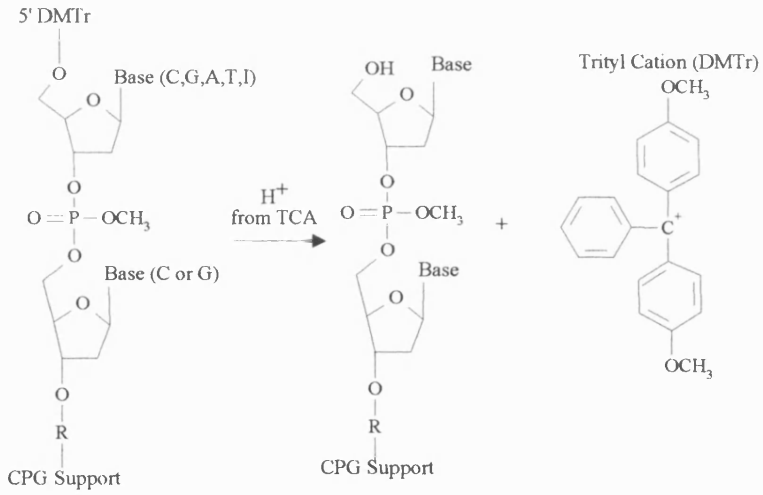
The four stages of oligonucleotide synthesis

A summary (Applied Biosystems manual) of the four stages of oligonucleotide synthesis, using automated solid-phase phosphotriester and phosphoramidite chemistry (Gait, 1984).

1) Detritylation

The purpose of this step is to remove the dimethoxytrityl protecting group from the 5′-OH of the nucleoside anchored to the column using TCA (Fig. 99).

Figure 99 Detritylation of the 5′ end nucleoside.



The 5′ hydroxyl is then available for further reaction.

2) Addition

Tetrazole is added to the column simultaeneously with the nucleoside derivative, so that it will protonate the nitrogen of the diisoprapyl group on the 3′ phosphorus. This makes the phosphoramidite susceptible to nucleophilic attack, by the detritylated hydroxyl of the growing DNA strand attached to the column. This results in a (5′ to 3′) internucleotide linkage *via* a trivalent phosphorus. (Fig. 100)

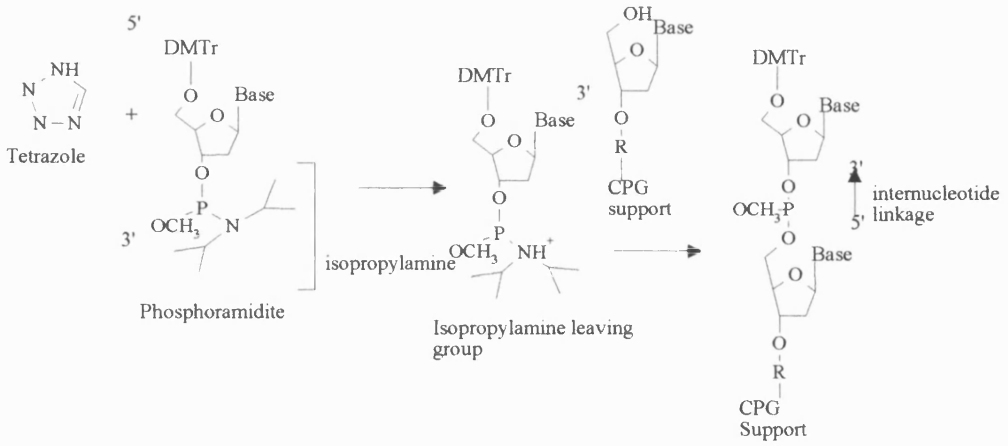


Figure 100 The coupling of a phosphoramidite to the nucleotide on the column.

3) Capping

Although the coupling reaction should have an efficiency of between 98-100%, a capping procedure is required to terminate any chains which did not undergo addition, thus limiting the lengths of any impurities and facilitating post-synthetic purification. The deoxynucleotides which reacted with the phosphoramidite previously, are blocked with a dimethoxytntyl group, whereas the unreacted chains have a free 5′OH that requires capping. This is achieved using an acetylating agent formed from equal volumes of equimolar acetic anhydride and dimethylaminopyridine (DMAP) (Fig. 101).

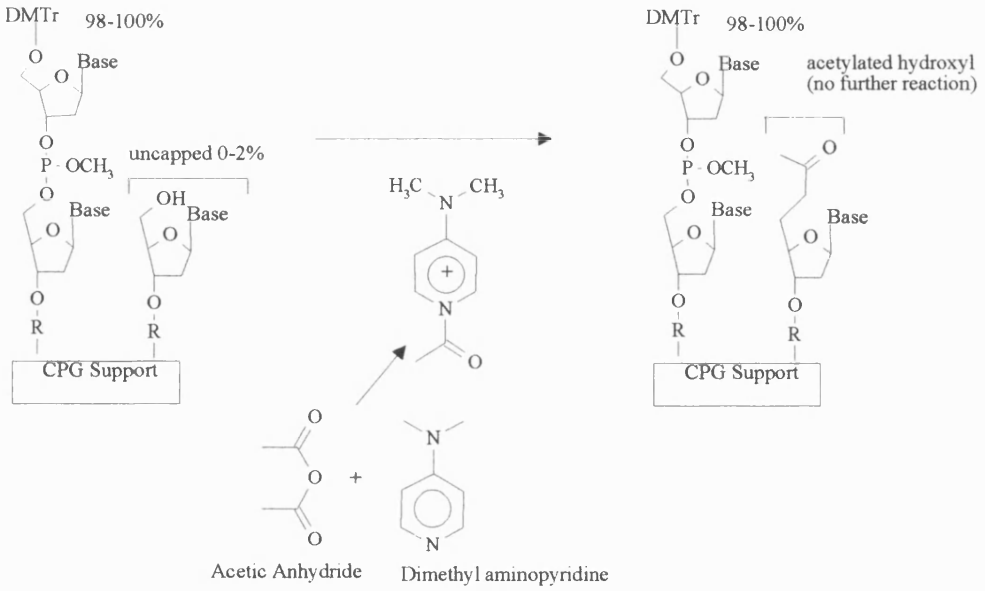


Figure 101 Capping reaction to terminate unreacted chains.

4) Oxidation

The linkage formed in the coupling reaction is an unstable phosphite (trivalent phosphorus) triester one, which is susceptible to acid and base cleavage. Conversion to a more stable pentavalent phosphate triester is therefore performed immediately after the capping step, hence also avoiding the water from the oxidising solution causing acetic anhydride to form acetic acid around the oligonucleotides. Iodine is the oxidising agent and water the oxygen donor in a basic lutidine/tetrahydroturan (THF) solution. On delivery to the column, an iodine lutidine complex is formed with the rivalent phosphorus, where it then undergoes displacement by water to a pentavalent phosphorous, within 30 seconds (Fig.102).

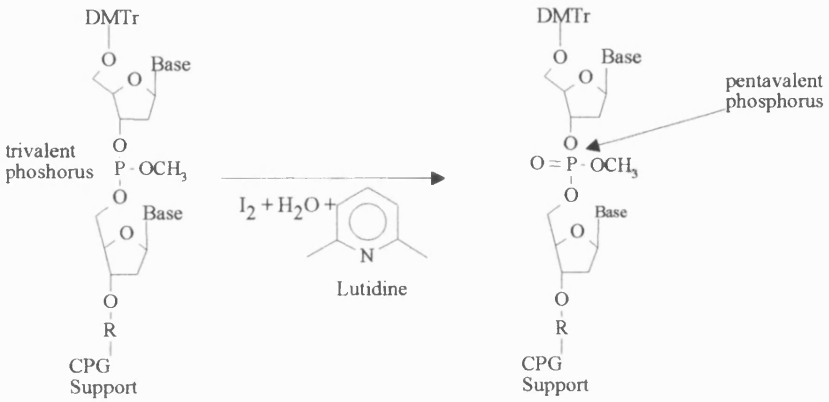


Figure 102 Oxidation of the internucleotide linkage to a stable pentavalent phosphorous.

These four steps were repeated for each nucleoside addition to the DNA strand, resulting in an oligonucleotide attached to the CPG column at the 3′ terminal and protected by a dimethoxytrityl group at the 5′ terminal.

APPENDIX III

The 400 MHz NOESY ^{1}H NMR spectrum of the DNA duplex 5′-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12}C^{13}G^{14})_2$

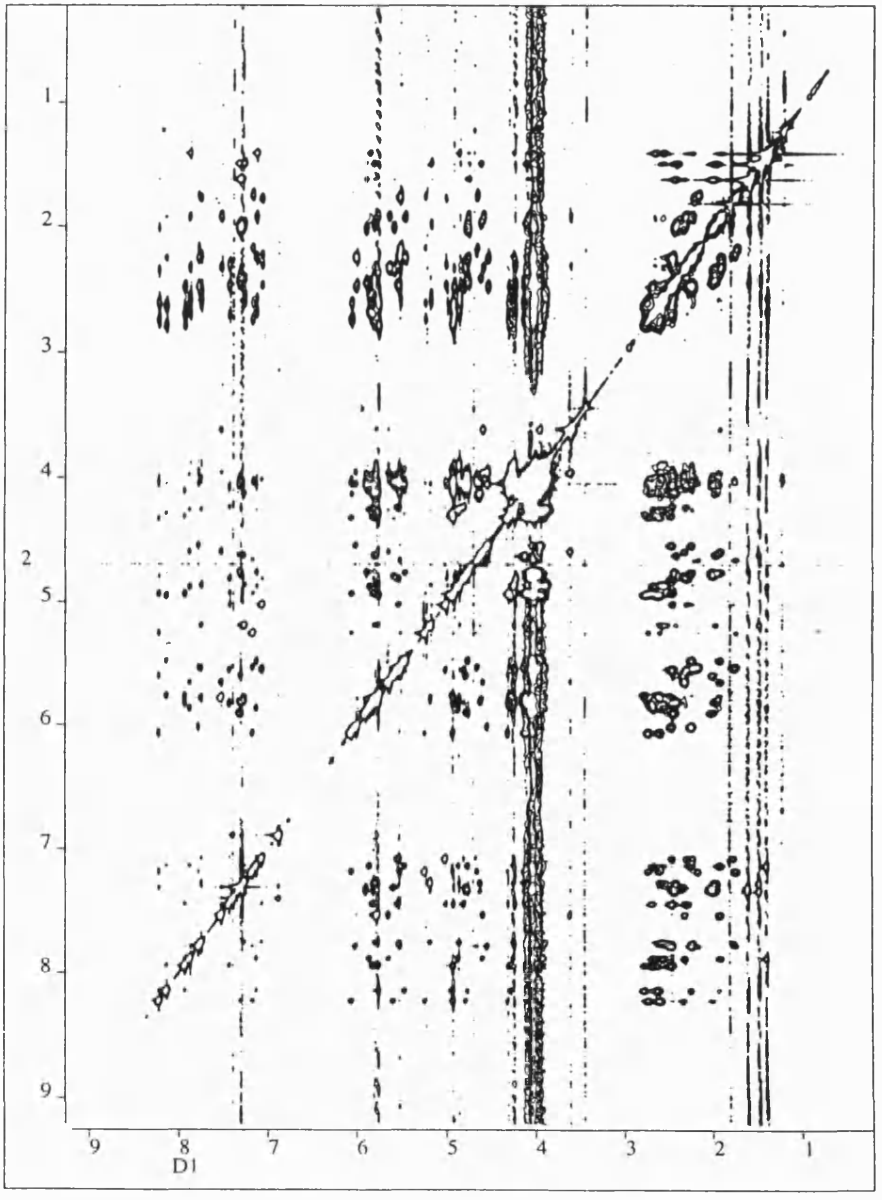


Figure 103 The 400 MHz NOESY ^{1}H NMR spectrum of the DNA duplex 5′-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12}C^{13}G^{14})_2$

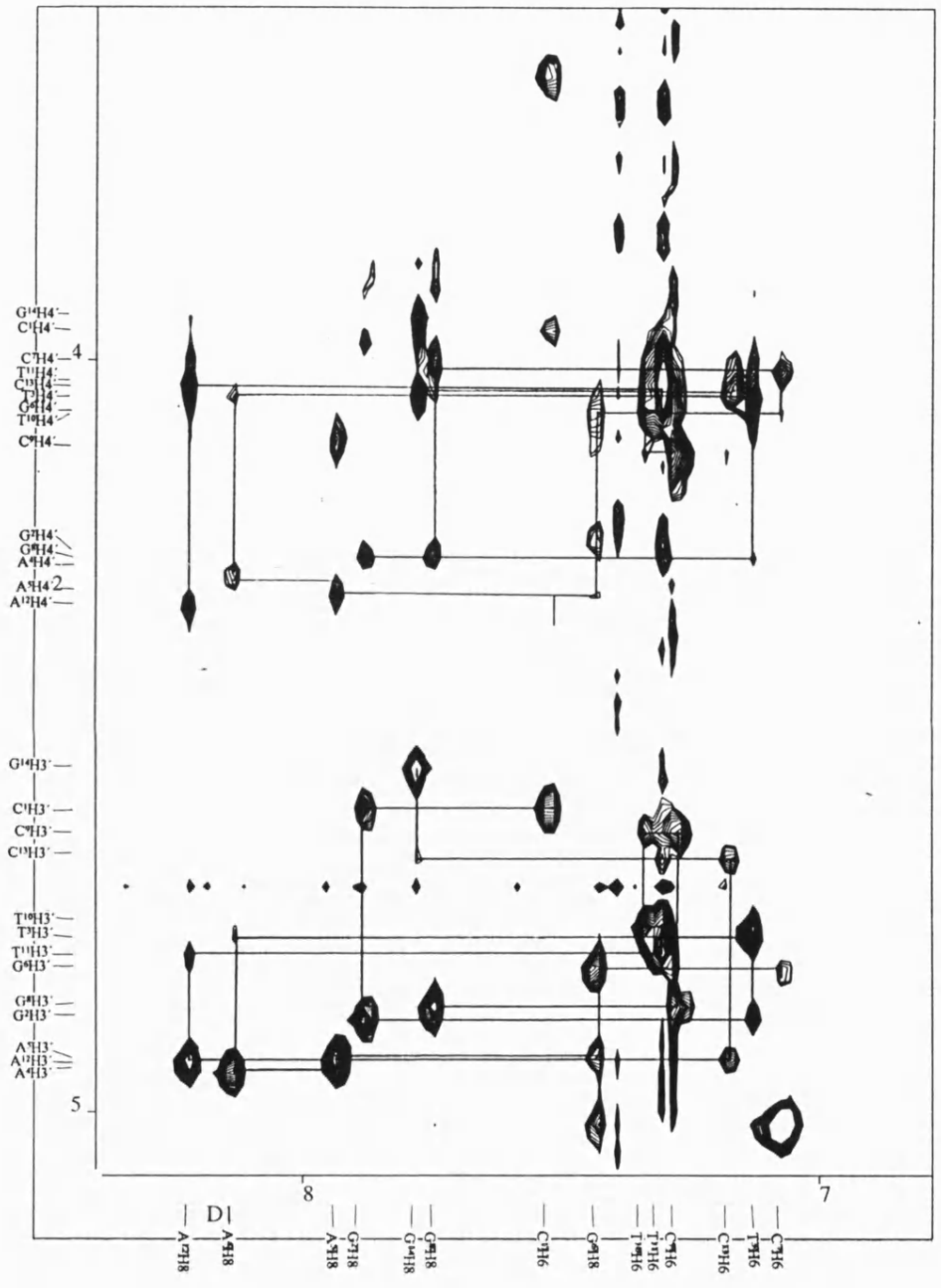


Figure 104 The 400 MHz NOESY ^{1}H NMR spectrum of the DNA duplex 5′-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12}C^{13}G^{14})_{2}$ The H6/H8→H3′ region.

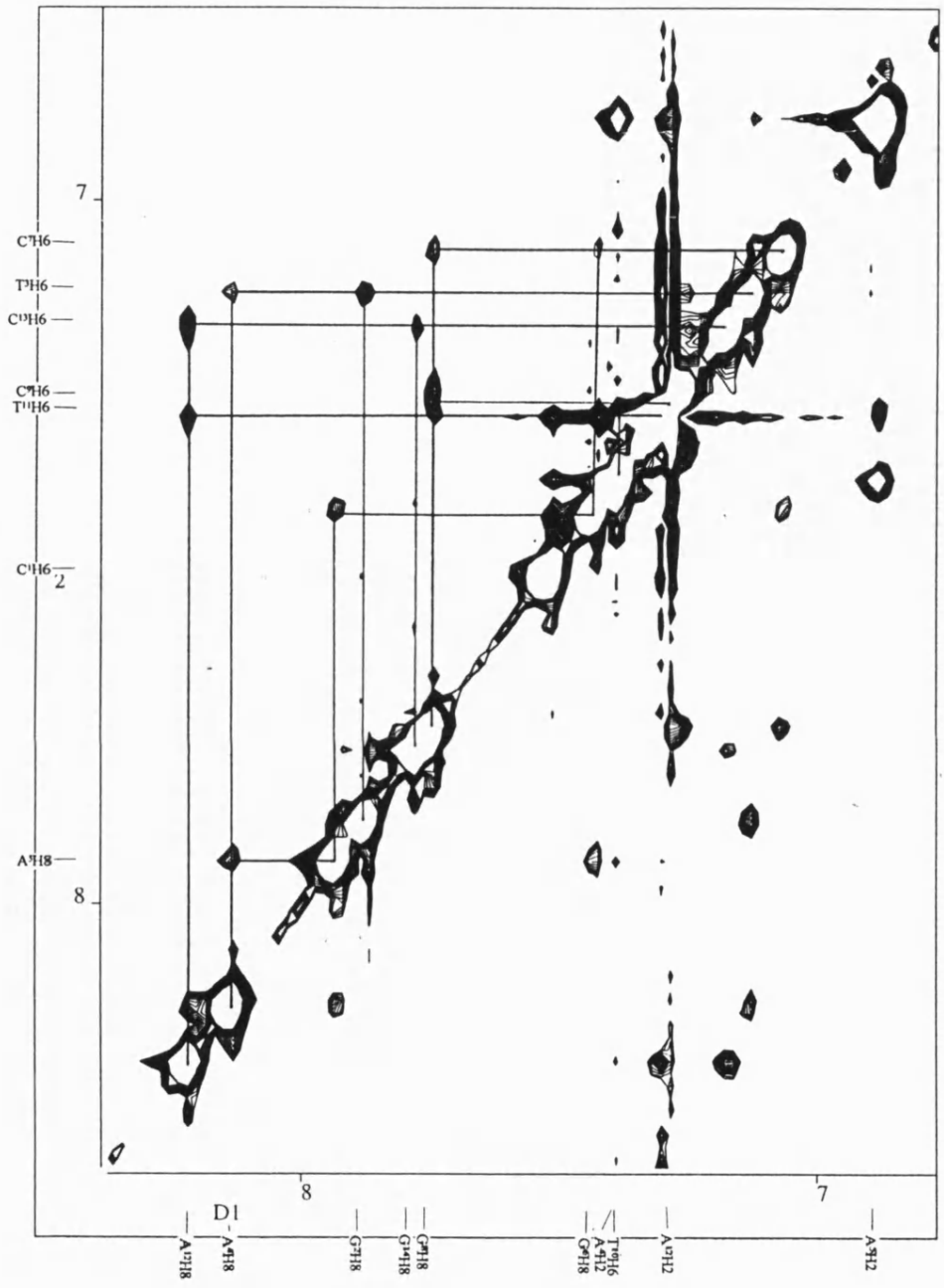


Figure 105 The 400 MHz NOESY ^{1}H NMR spectrum of the DNA duplex 5′-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12}C^{13}G^{14})_{2}$. The H6/H8→H6/H8 region.

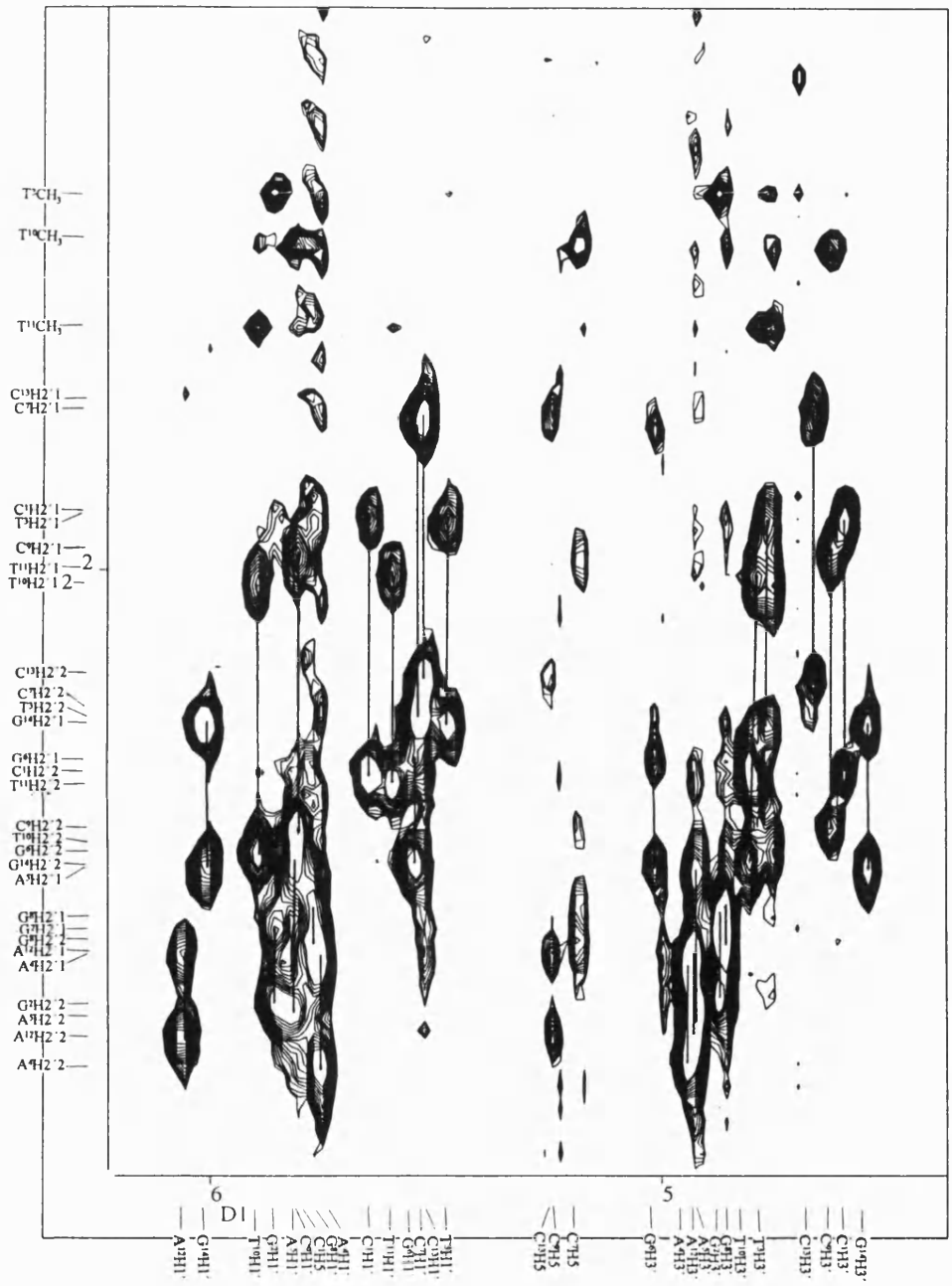


Figure 106 The 400 MHz NOESY ^{1}H NMR spectrum of the DNA duplex 5'-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12}C^{13}G^{14})_{2}$. The H1'→H2'1/H2'2 region.

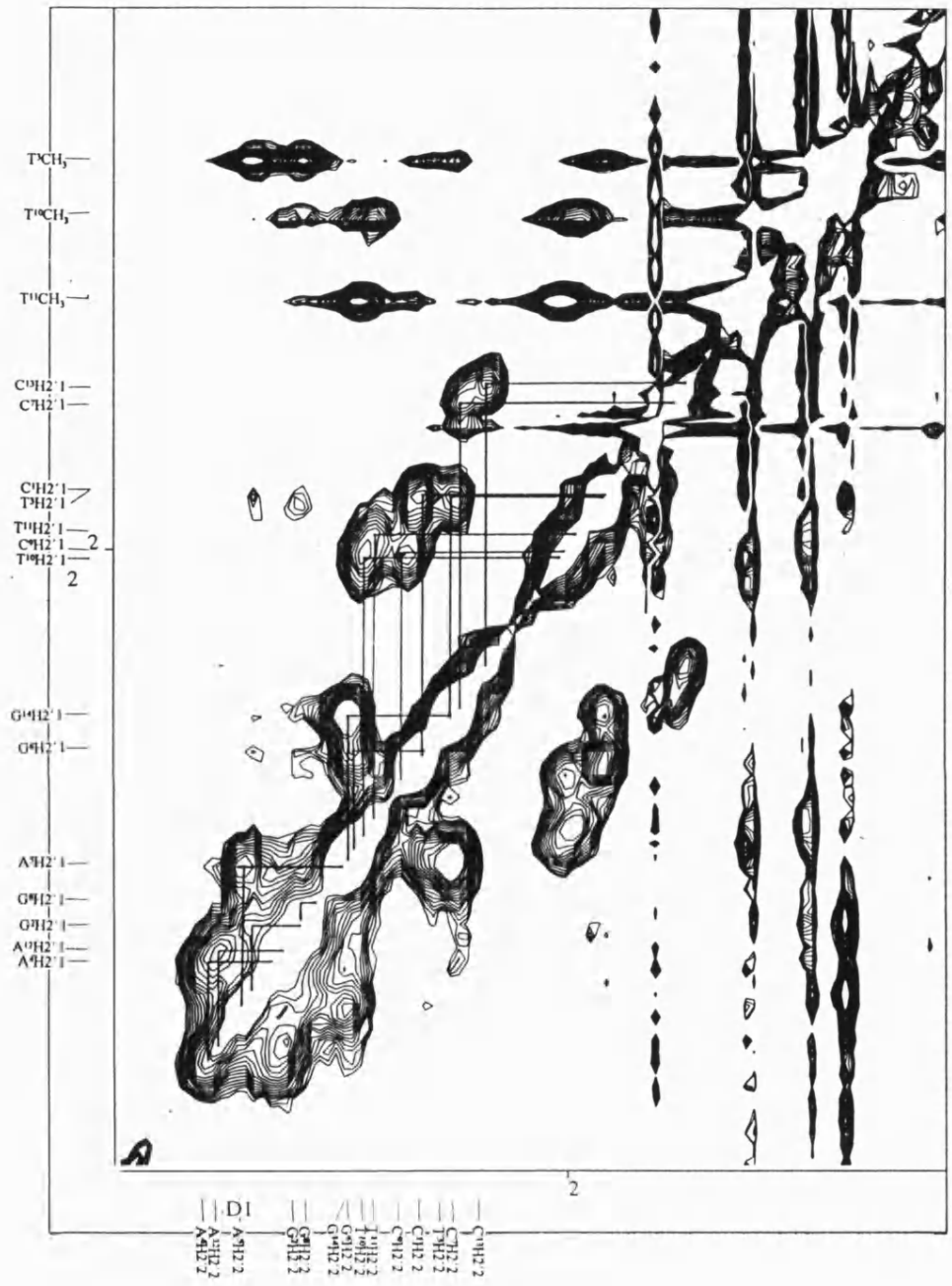


Figure 107 The 400 MHz NOESY ^{1}H NMR spectrum of the DNA duplex 5′-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12}C^{13}G^{14})_{2}$. The H2′1→H2′2 region.

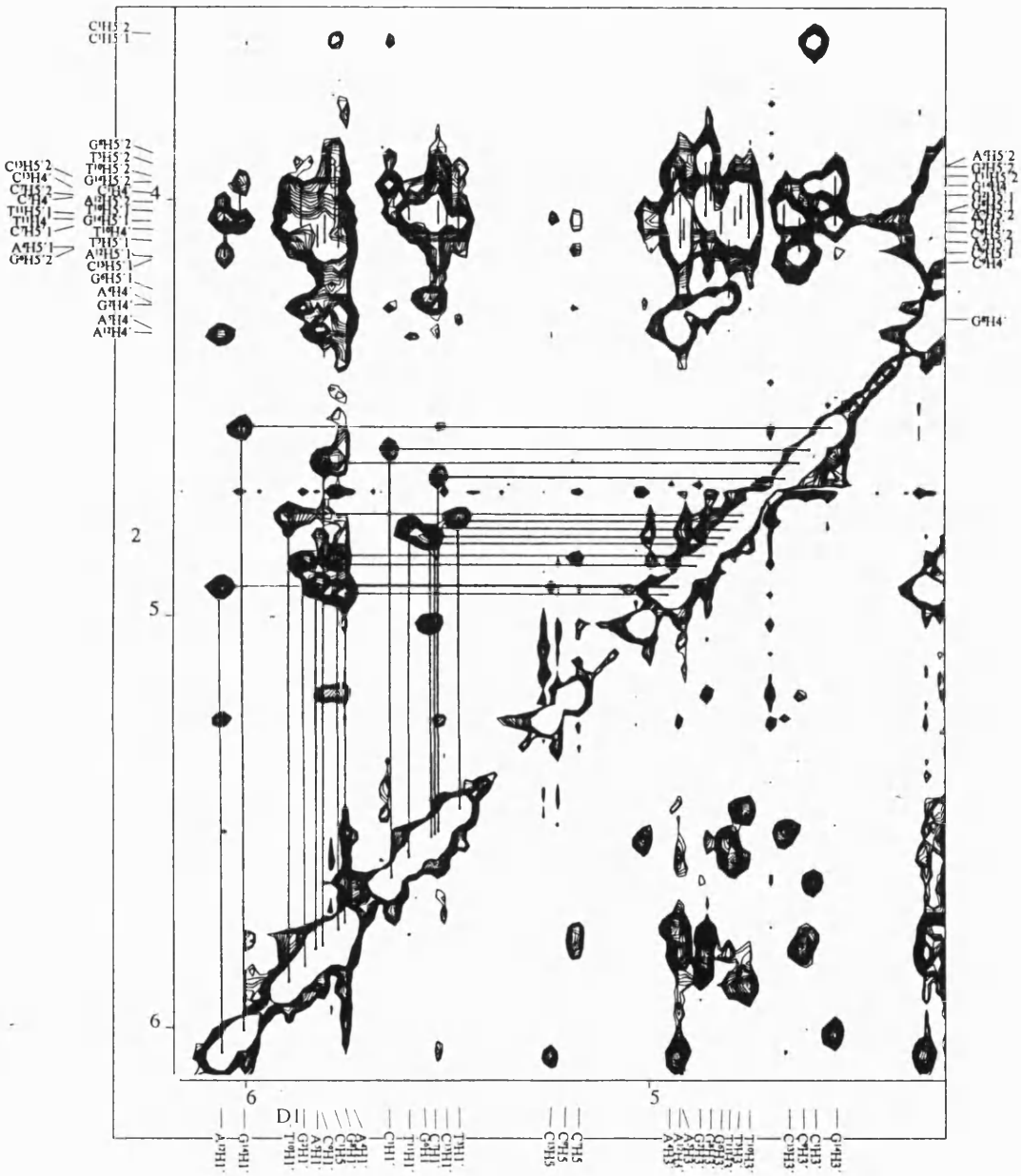


Figure 108 The 400 MHz NOESY ^{1}H NMR spectrum of the DNA duplex 5′-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12}C^{13}G^{14})_{2}$. The H1′/H3′→H4′/H5′1/H5′2 region.

APPENDIX IV

One-dimensional NMR spectrum of the DNA and adozelesin-DNA reaction mixture.

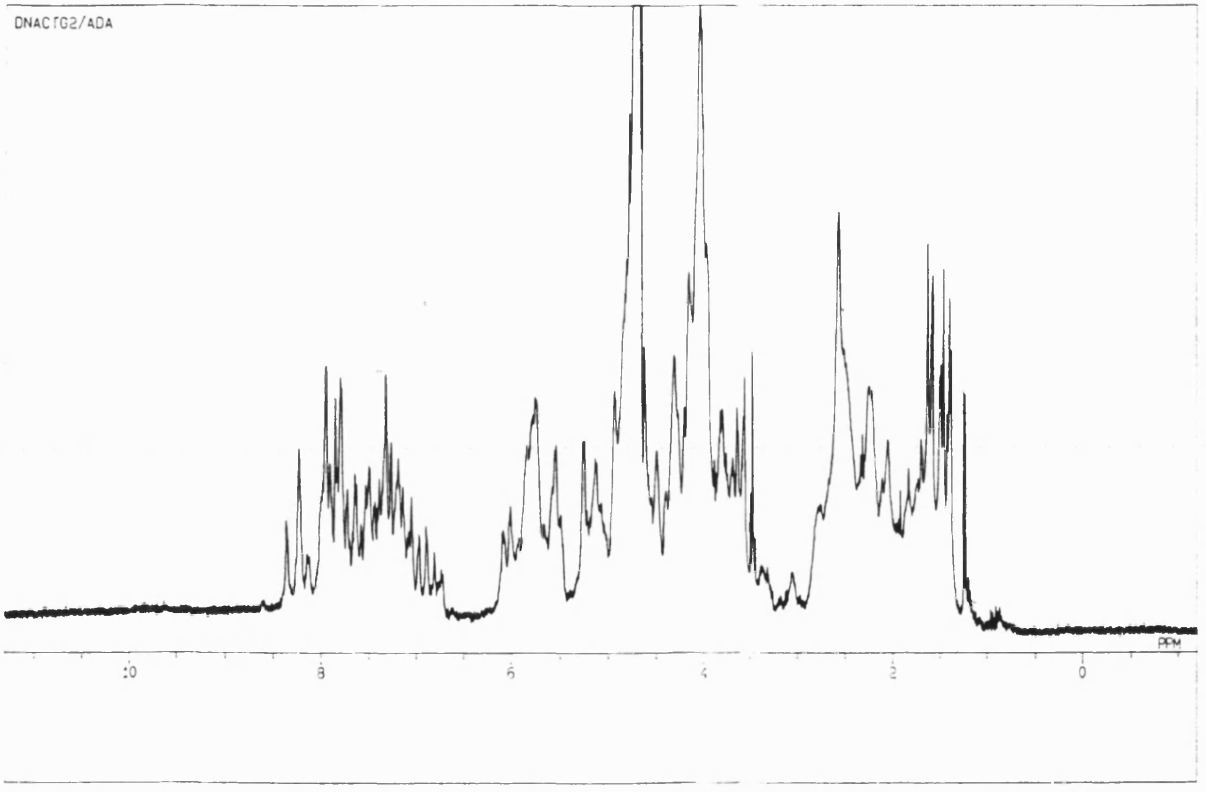


Figure 109 One-dimensional ^{1}H NMR spectrum of the 5′-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12}C^{13}G^{14})_2$ DNA duplex and adozelesin-DNA adduct reaction mixture.

APPENDIX V

The 400 MHz DQF-COSY spectrum of the adozelesin-5′-($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct.

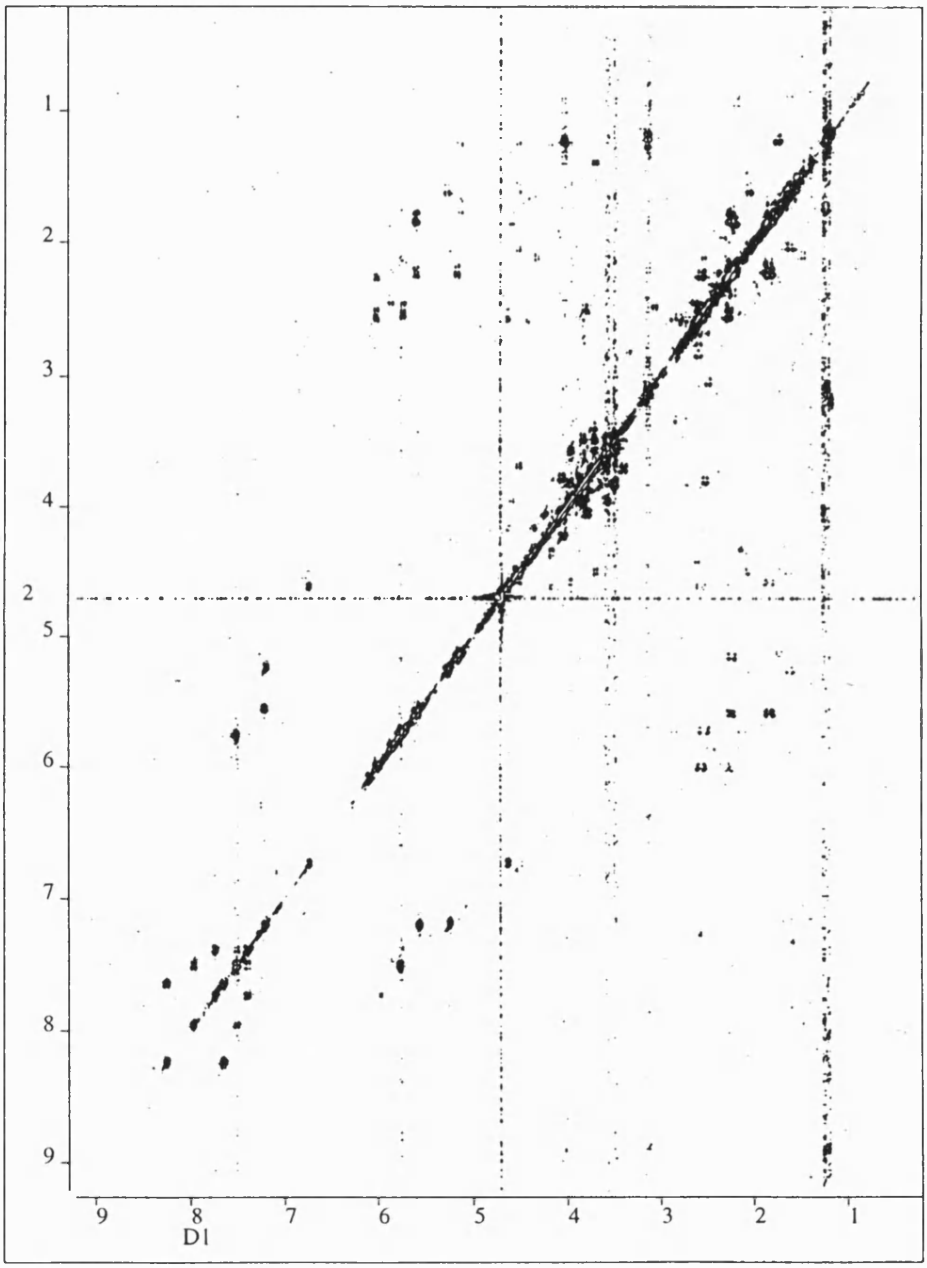


Figure 110 The 400 MHz DQF-COSY NMR spectrum of the adozelesin-5′-($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct.

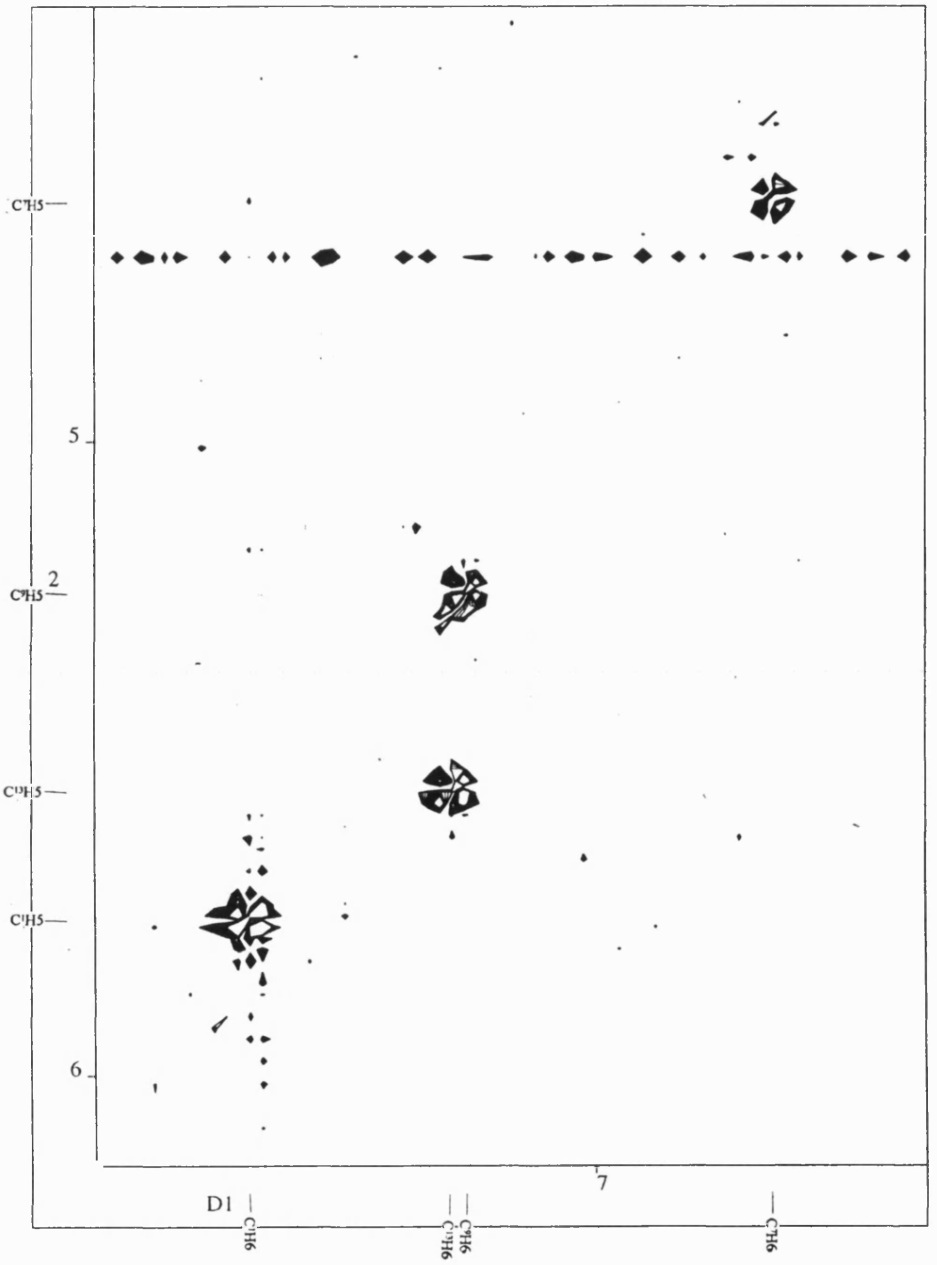


Figure 111 The 400 MHz DQF-COSY NMR spectrum of the adozelesin-5′-($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The CH5→CH6 region.

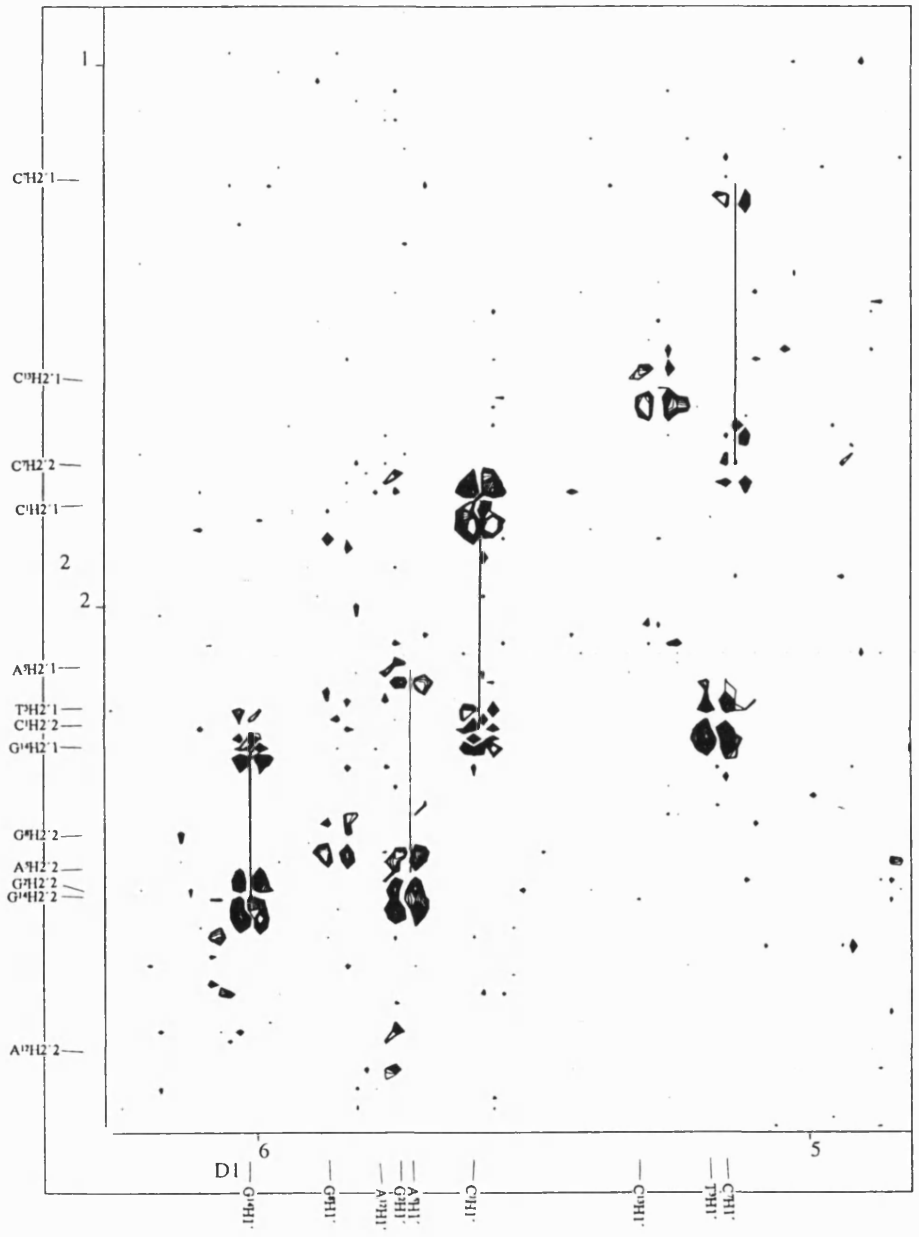


Figure 112 The 400 MHz DQF-COSY NMR spectrum of the adozelesin-5′-
$(C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12*}C^{13}G^{14})_2$ DNA adduct. The H1′→H2′1/H2′2 region.

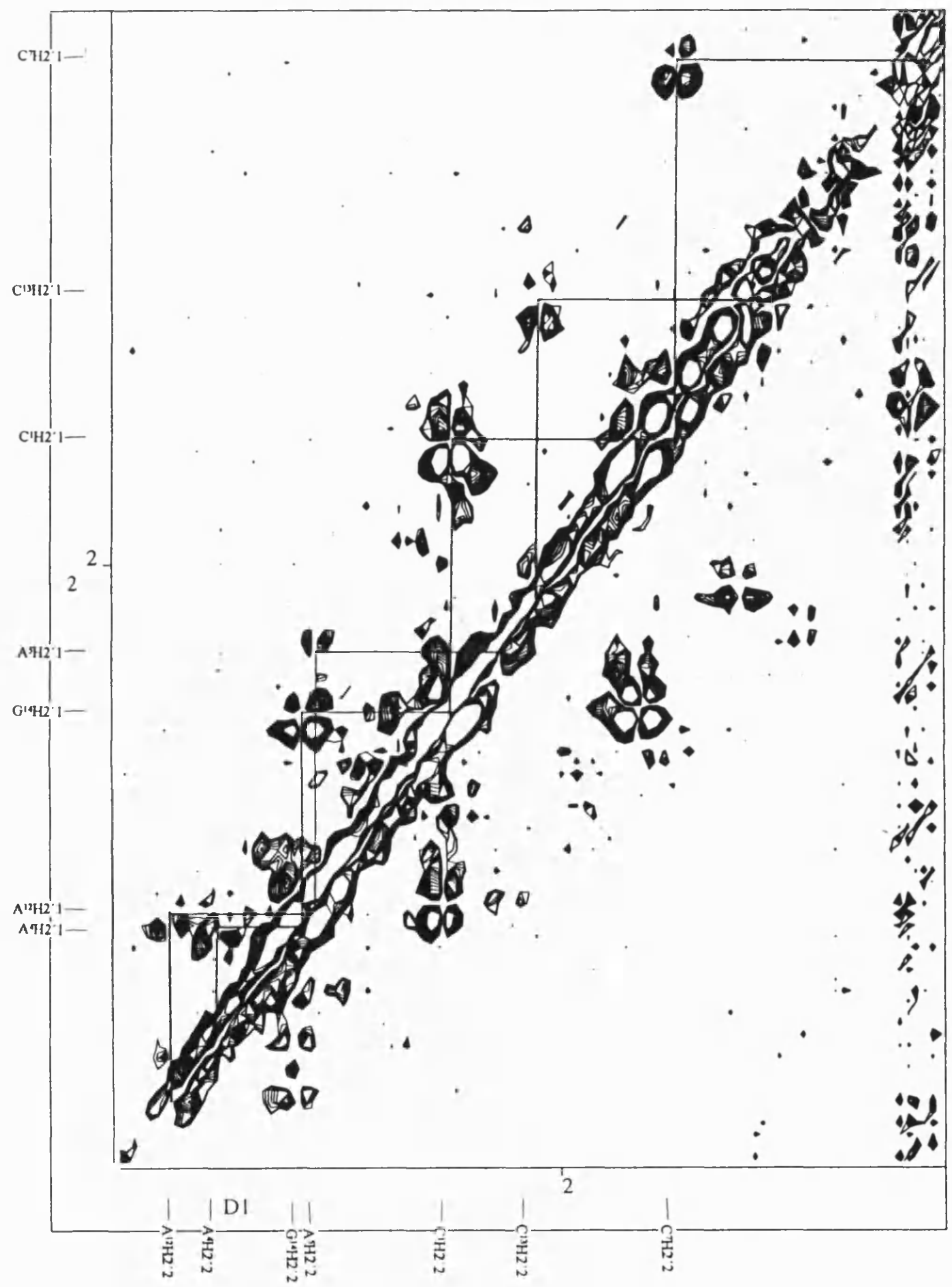


Figure 113 The 400 MHz DQF-COSY NMR spectrum of the adozelesin-5′-
$(C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12*}C^{13}G^{14})_2$ DNA adduct. The H2′1→H2′2 region.

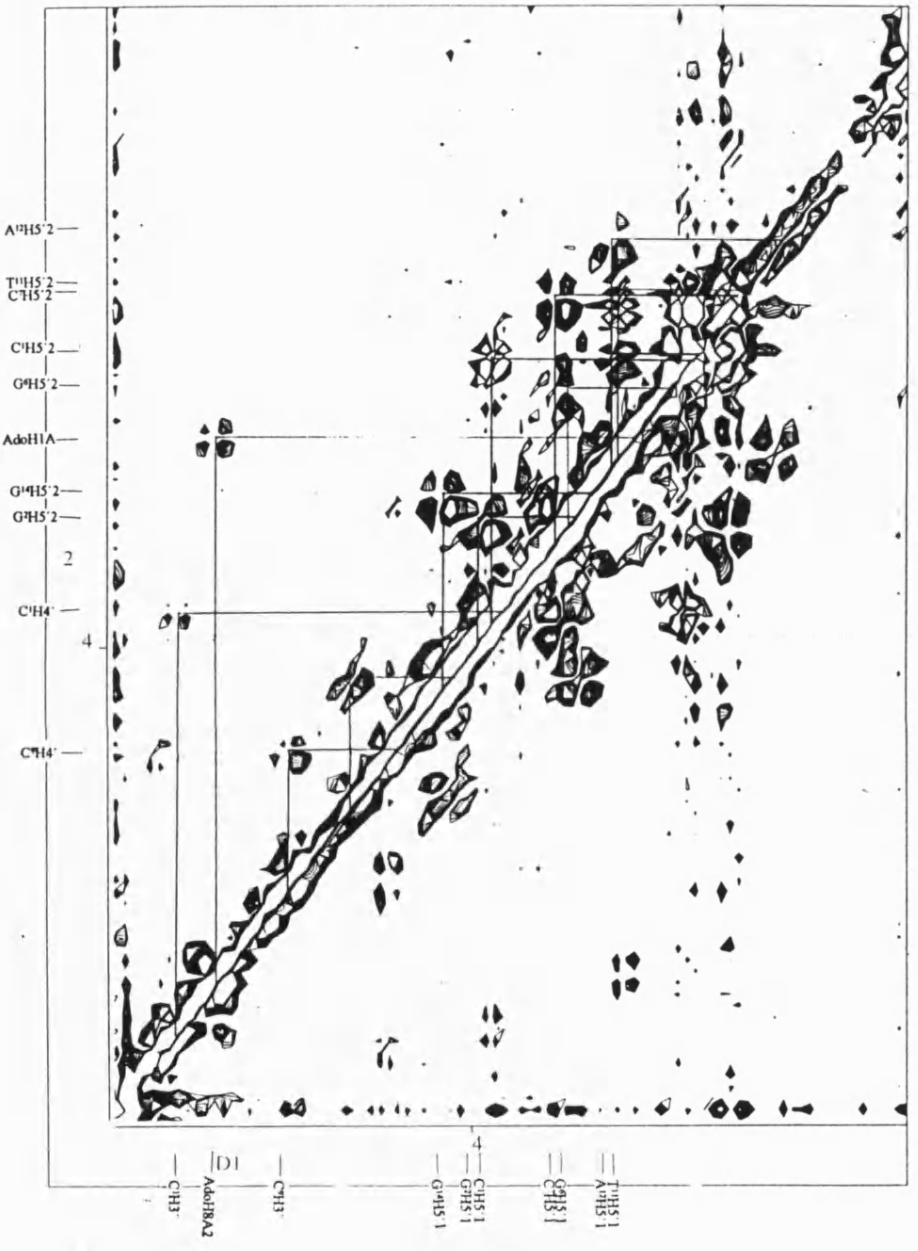


Figure 114 The 400 MHz DQF-COSY NMR spectrum of the adozelesin-5′-
$(C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12*}C^{13}G^{14})_2$ DNA adduct. The H3′, H4′, H5′1 and H5′2 region.

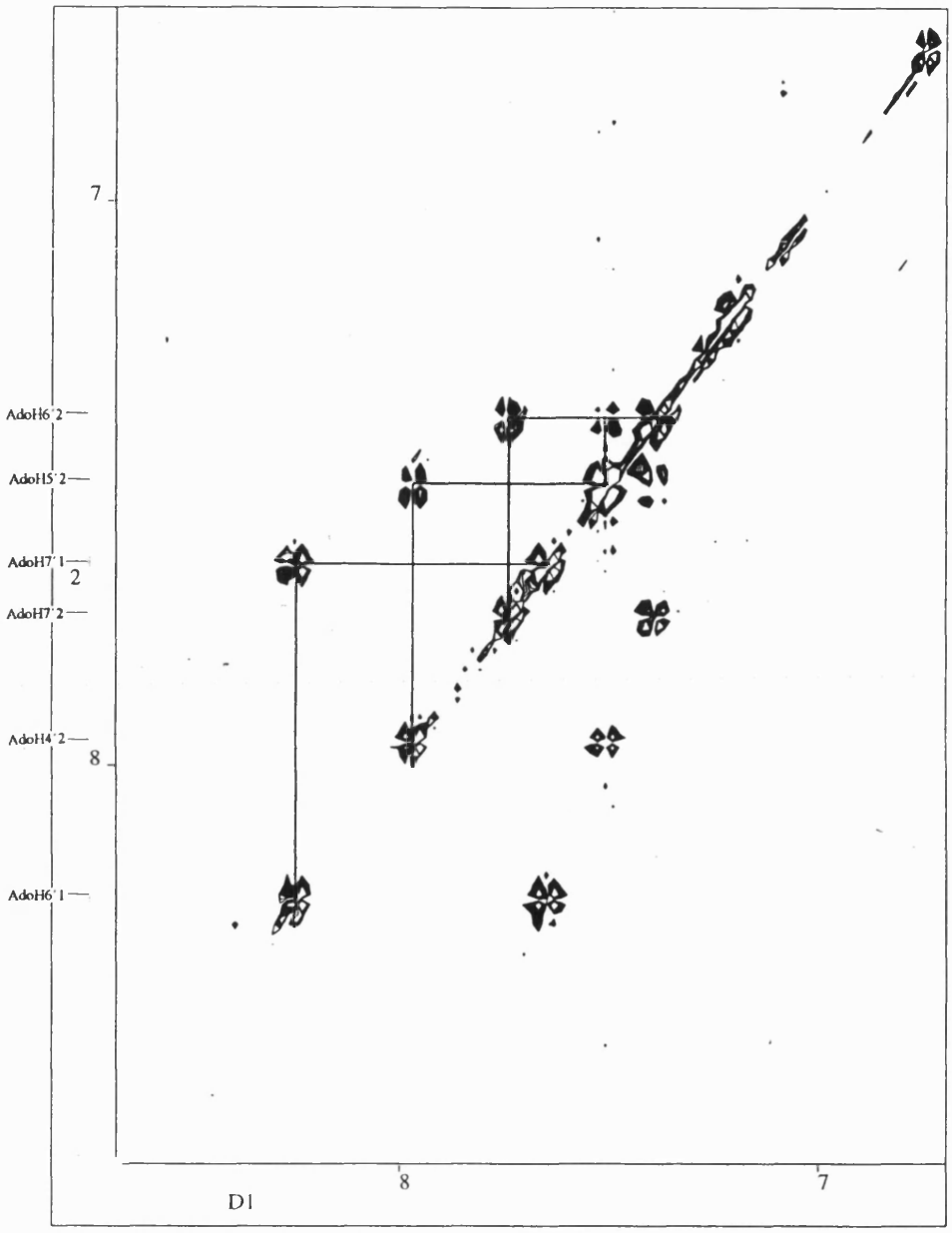


Figure 115 The 400 MHz DQF-COSY NMR spectrum of the adozelesin-5′-
$(C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12*}C^{13}G^{14})_2$ DNA adduct. The aromatics→aromatics region.

APPENDIX VI

The 400 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct.

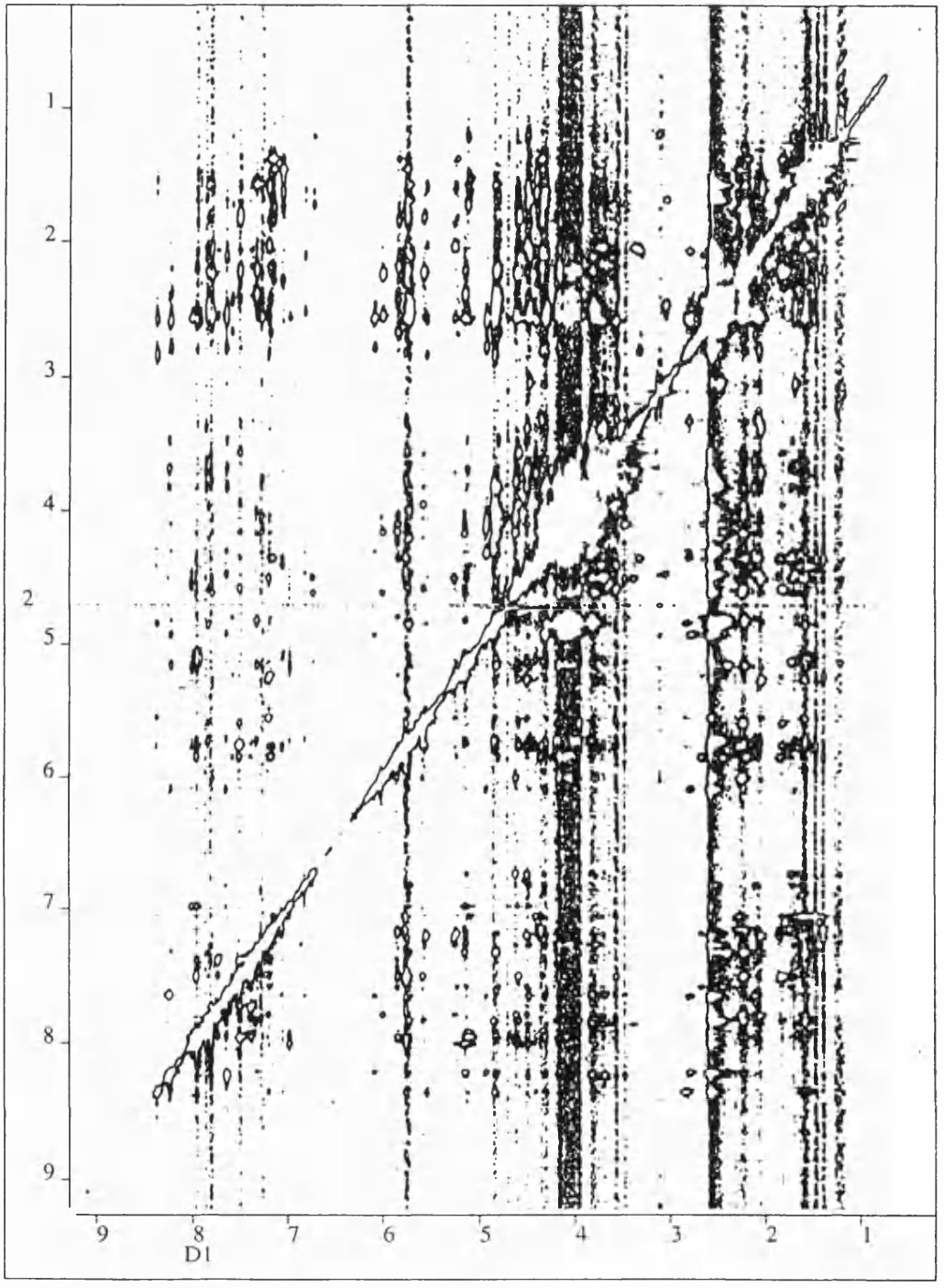


Figure 116 The 400 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct.

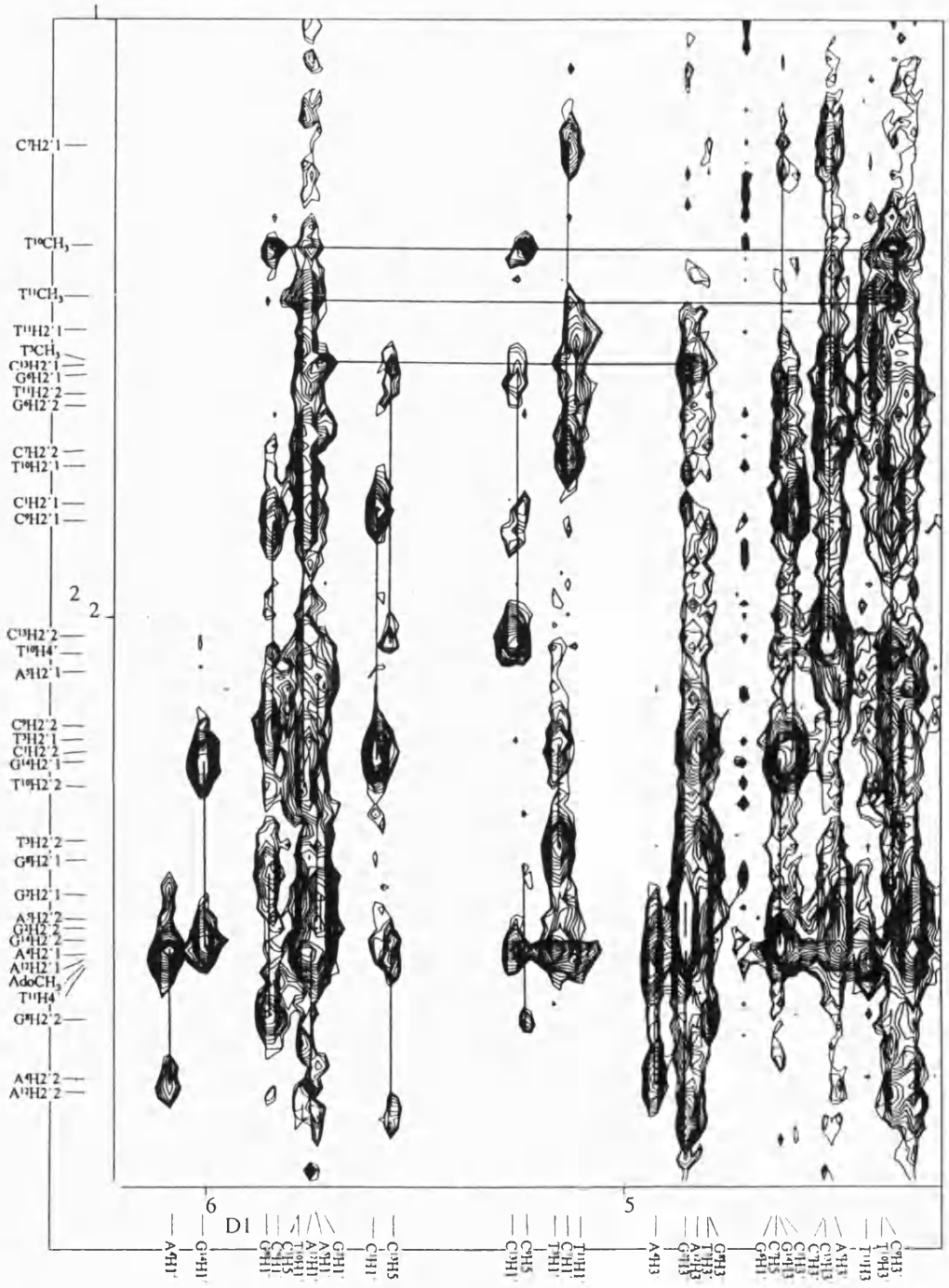


Figure 117 The 400 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The H1′→H2′1/H2′2 region.

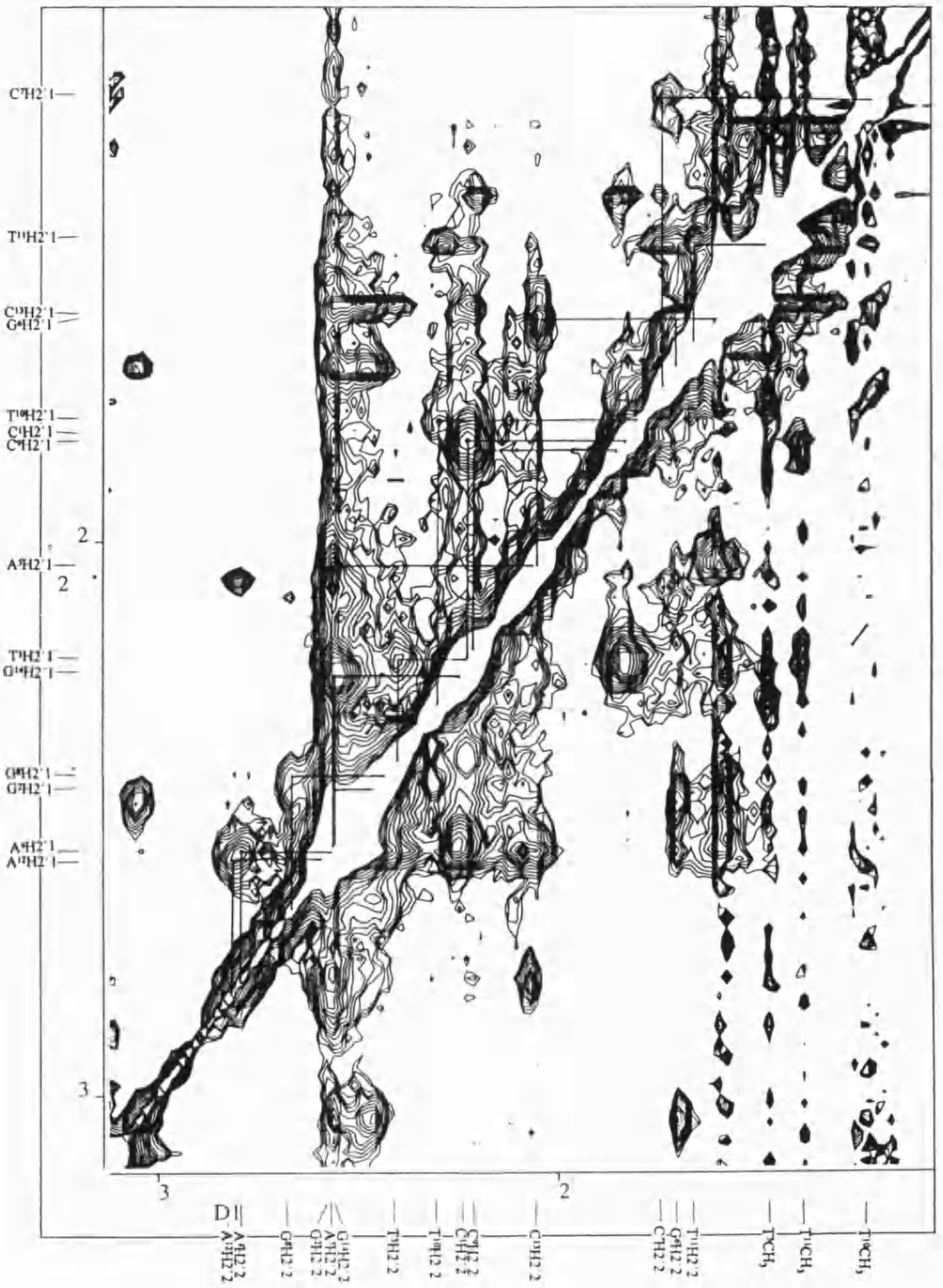


Figure 118 The 400 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The H2′1→H2′2 region.

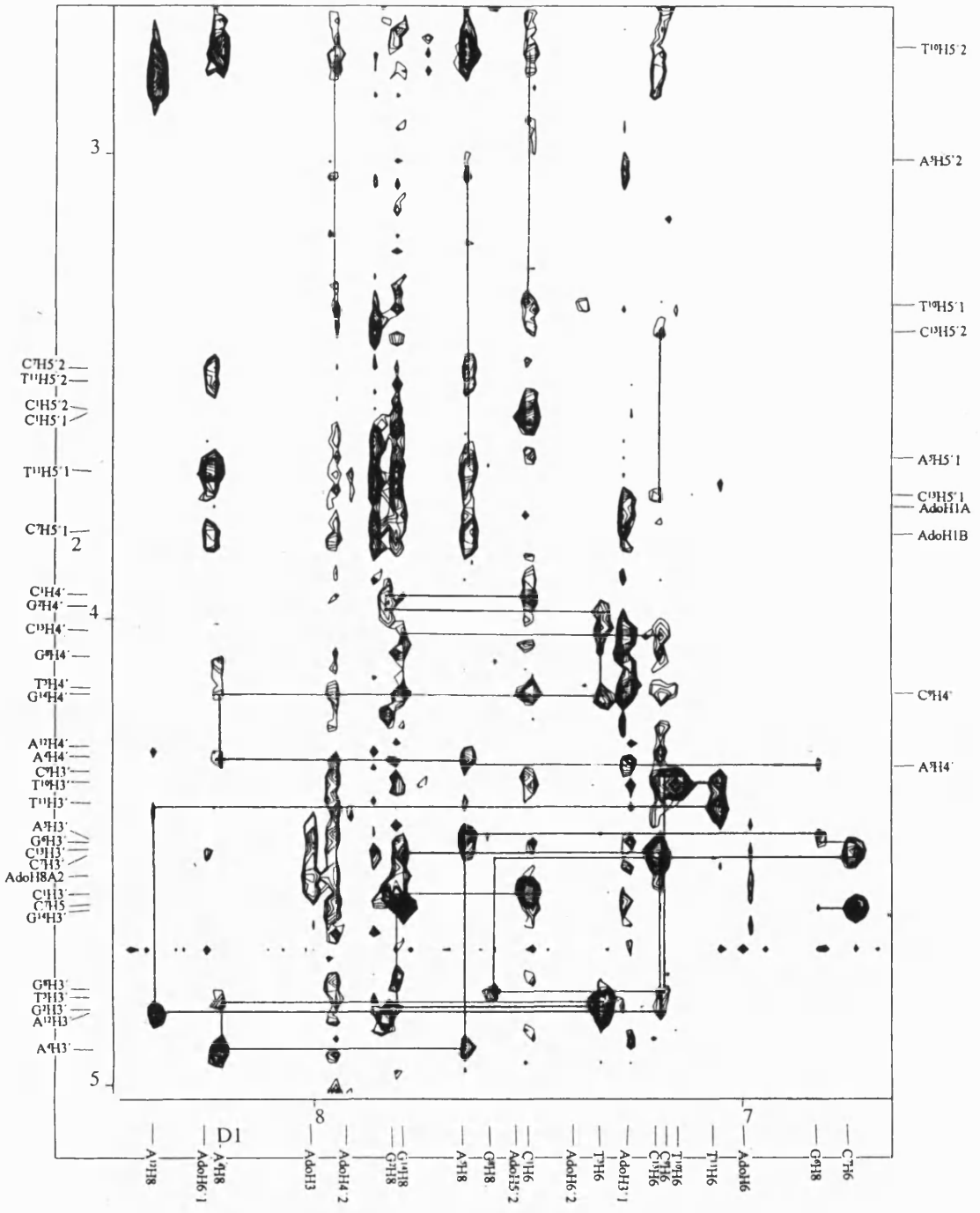


Figure 119 The 400 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-
($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12•}C^{13}G^{14}$)$_2$ DNA adduct. The aromatics→H3′/H4′/H5′1/H5′2 region.

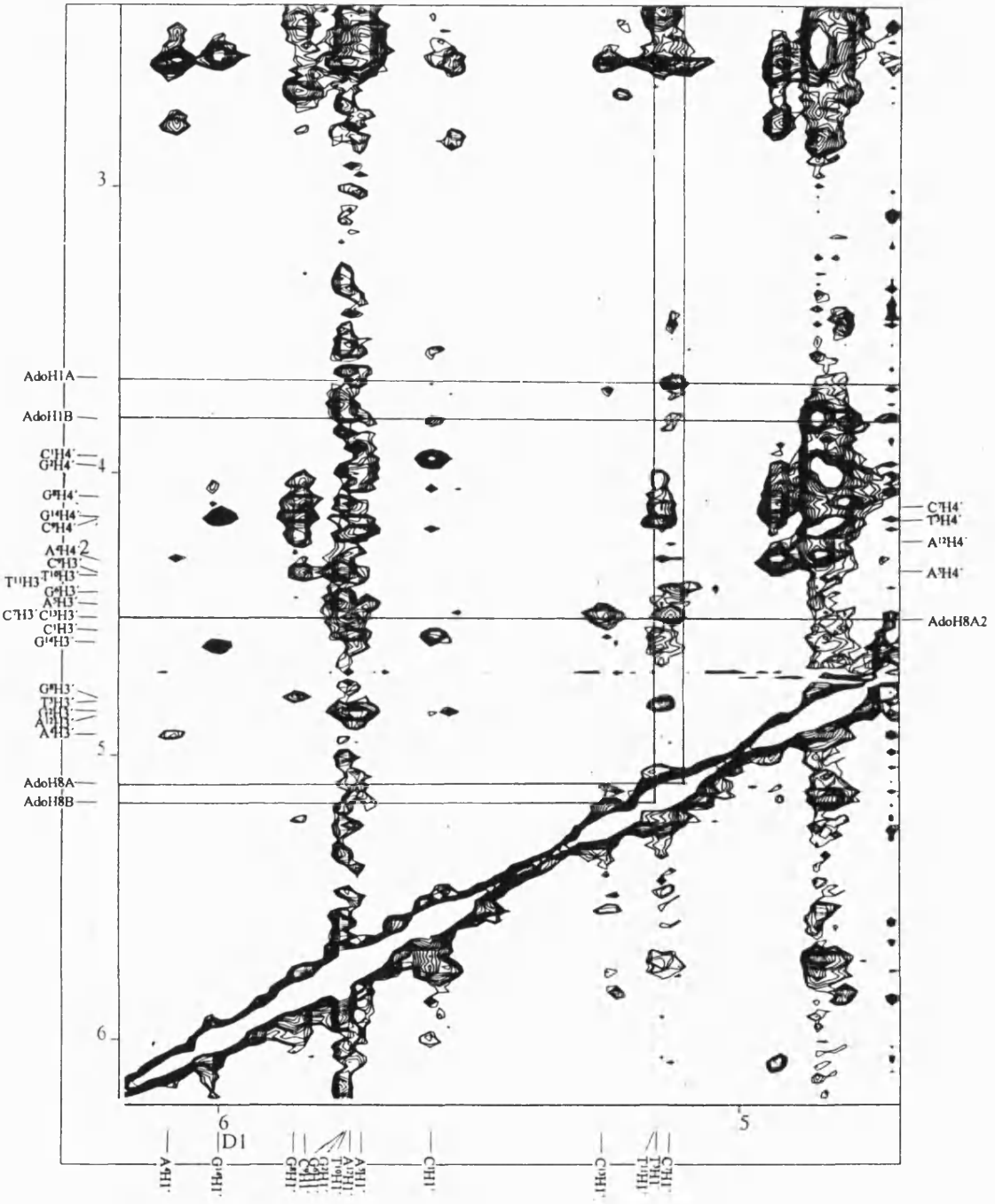


Figure 120 The 400 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The H1′→H3′/H4′ region.

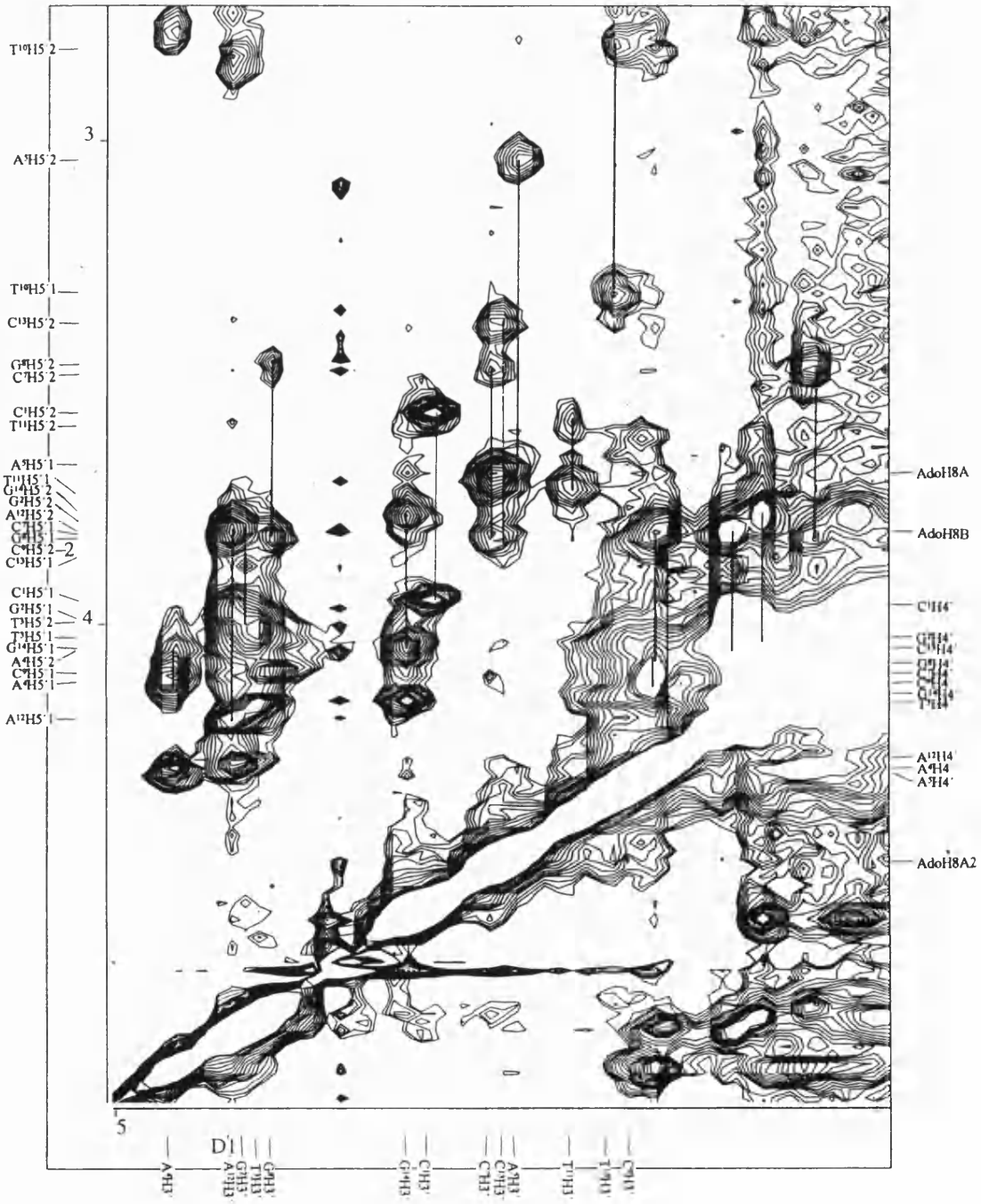


Figure 121 The 400 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The H3′→H4′/H5′1/H5′2 region.

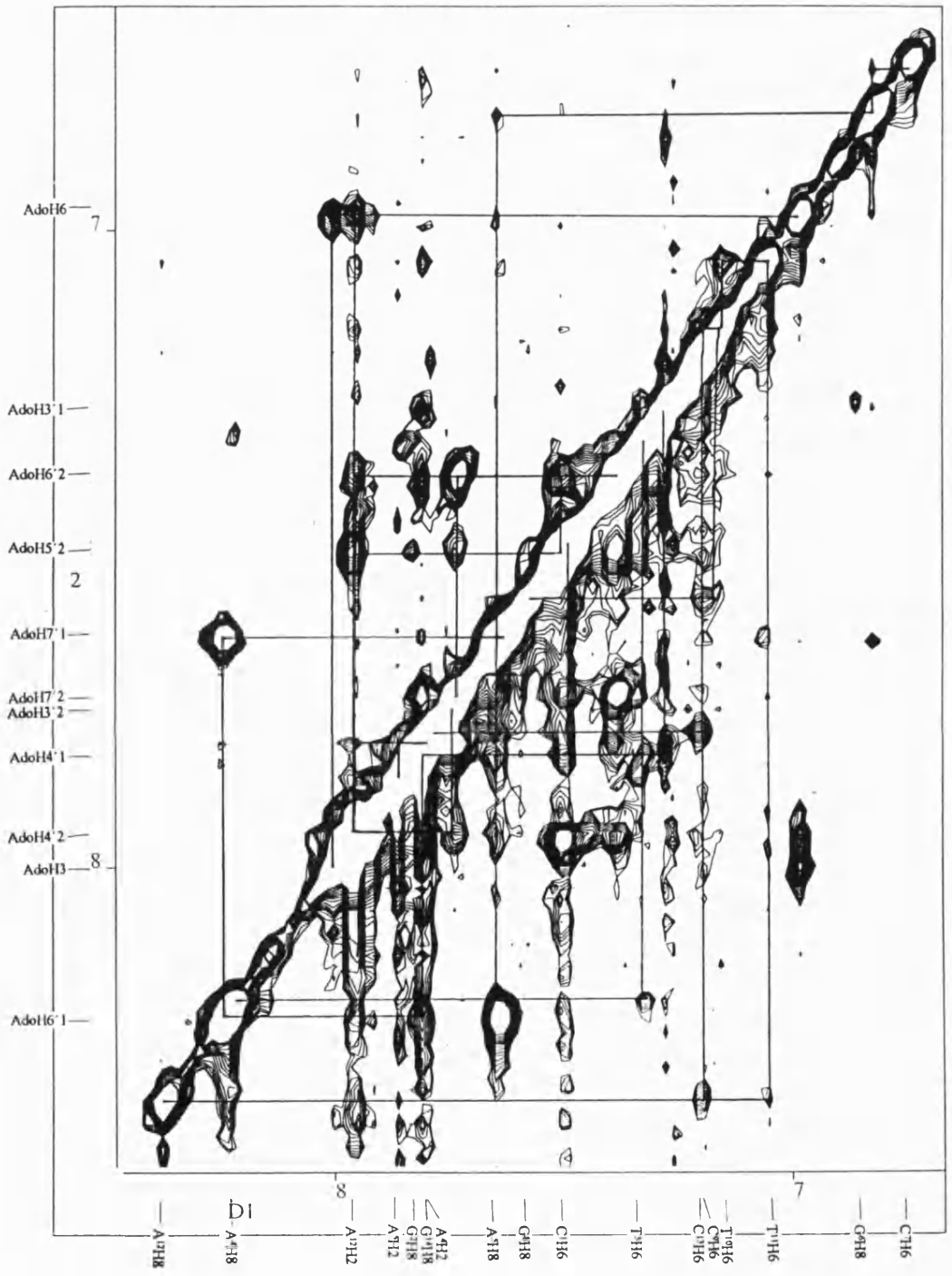


Figure 122 The 400 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The aromatics→aromatics region.

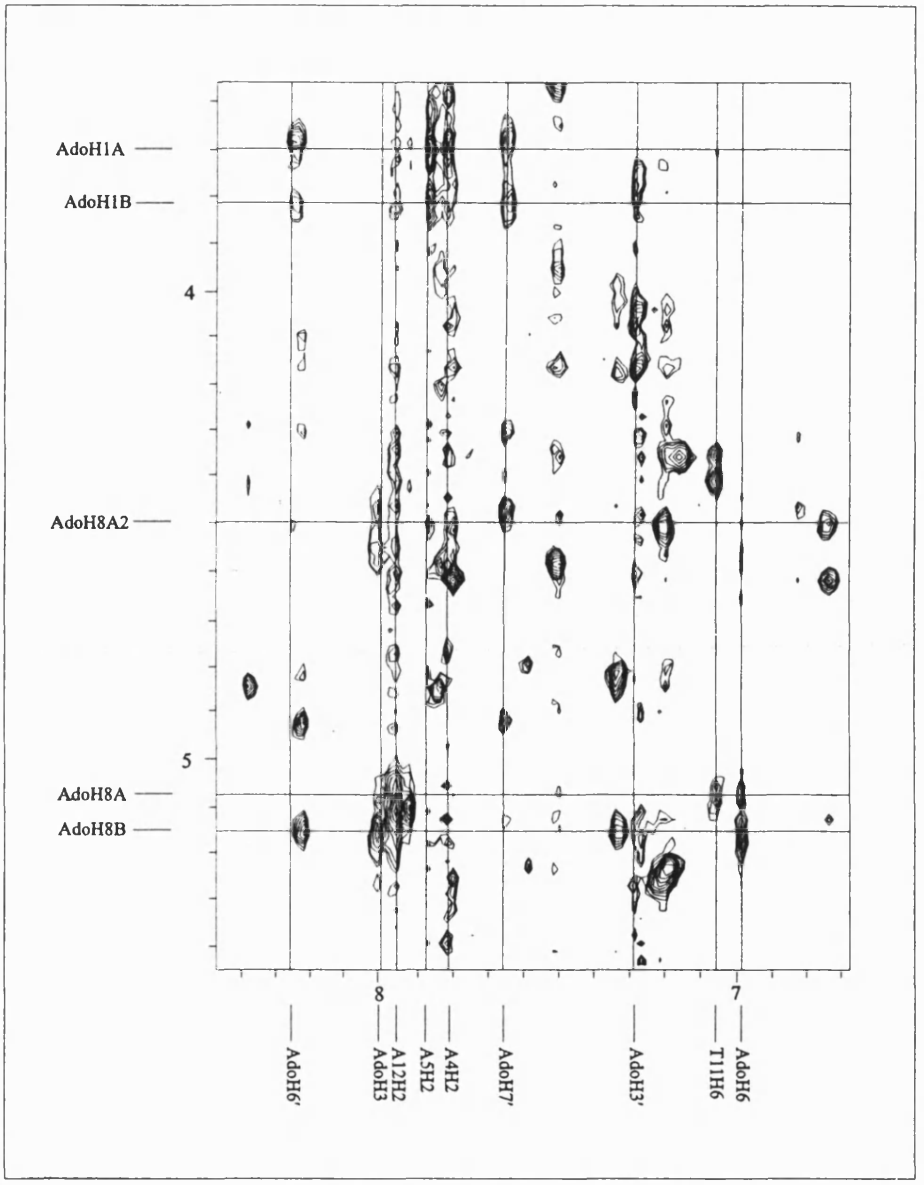


Figure 123 The 400 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. Adozelesin NOEs from the CPI headunit.

APPENDIX VII

The 600 MHz DQF-COSY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct.

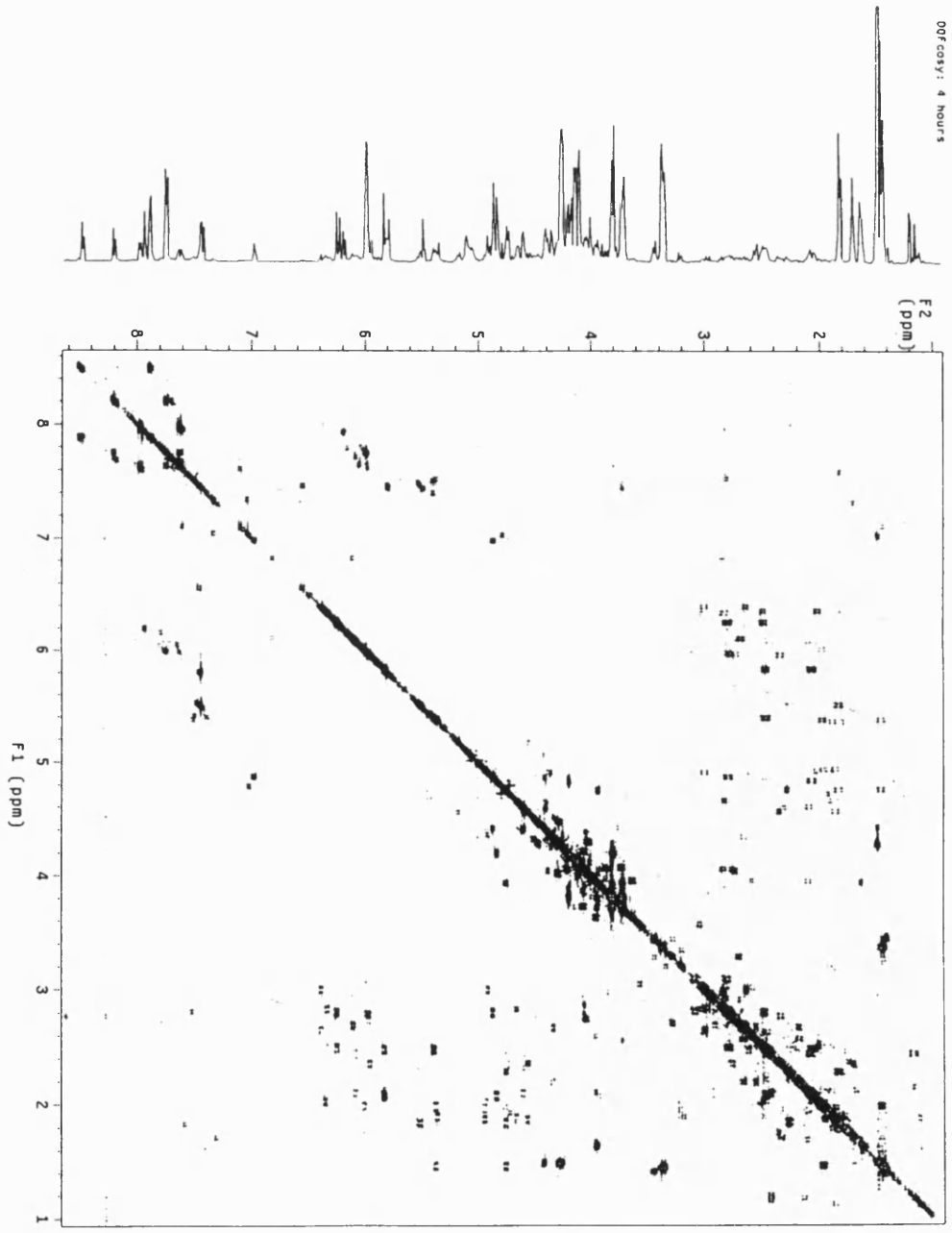


Figure 124 The 600 MHz DQF-COSY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct.

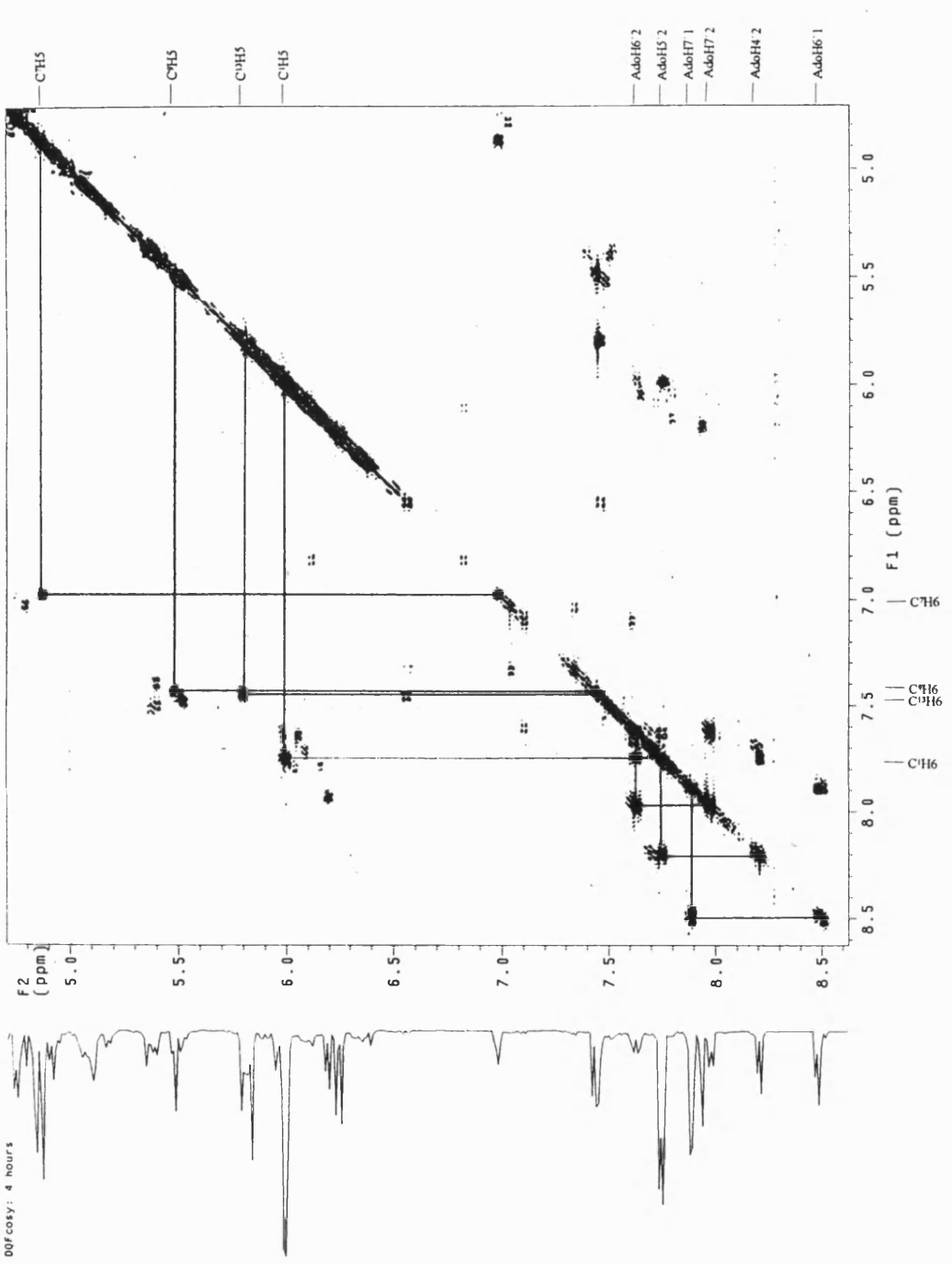


Figure 125 The 600 MHz DQF-COSY ^{1}H NMR spectrum of the adozelesin-5′-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14})_{2}$ DNA adduct. The aromatics→aromatics region and the CH6→CH6 region.

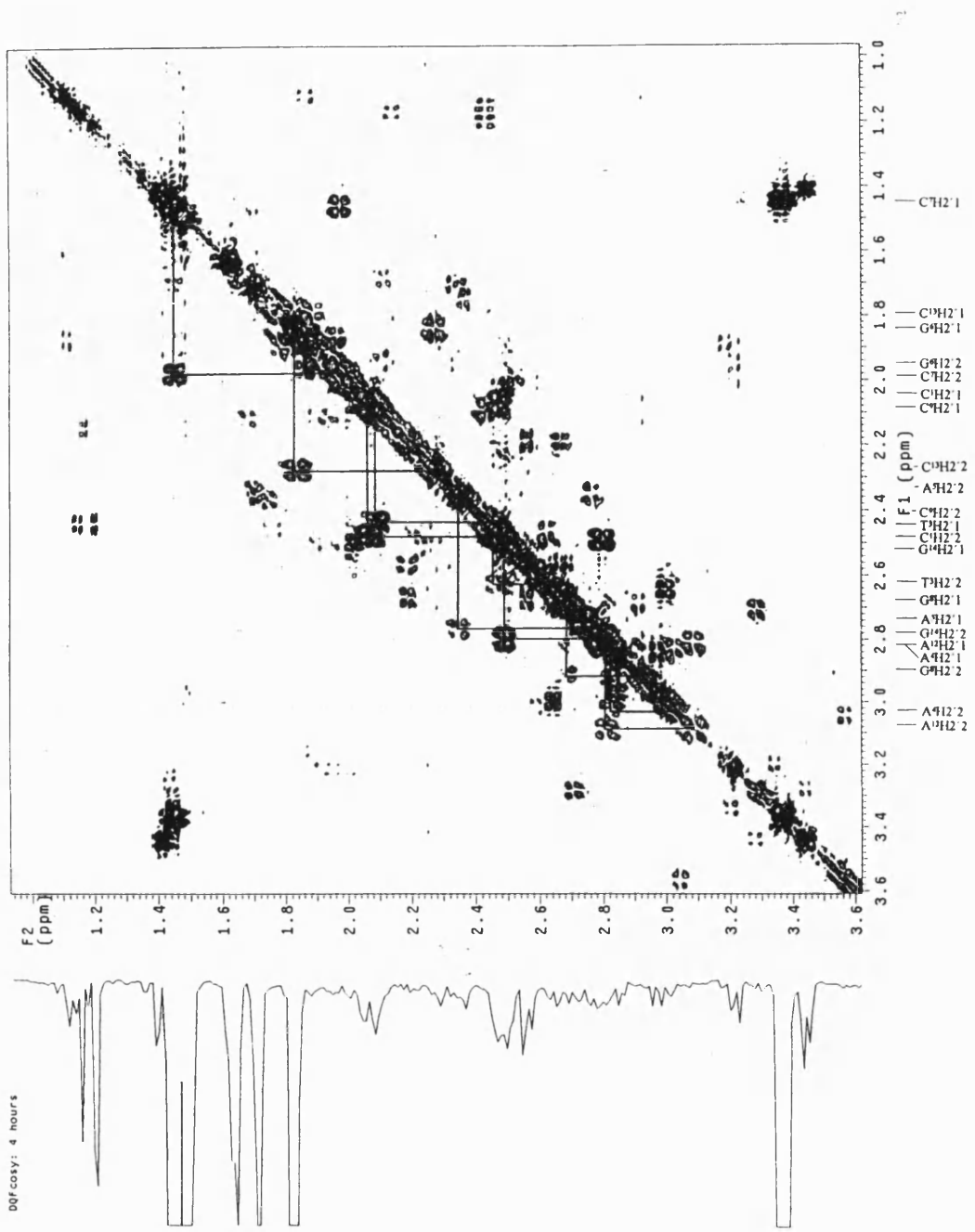


Figure 126 The 600 MHz DQF-COSY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The H2′1→H2′2.

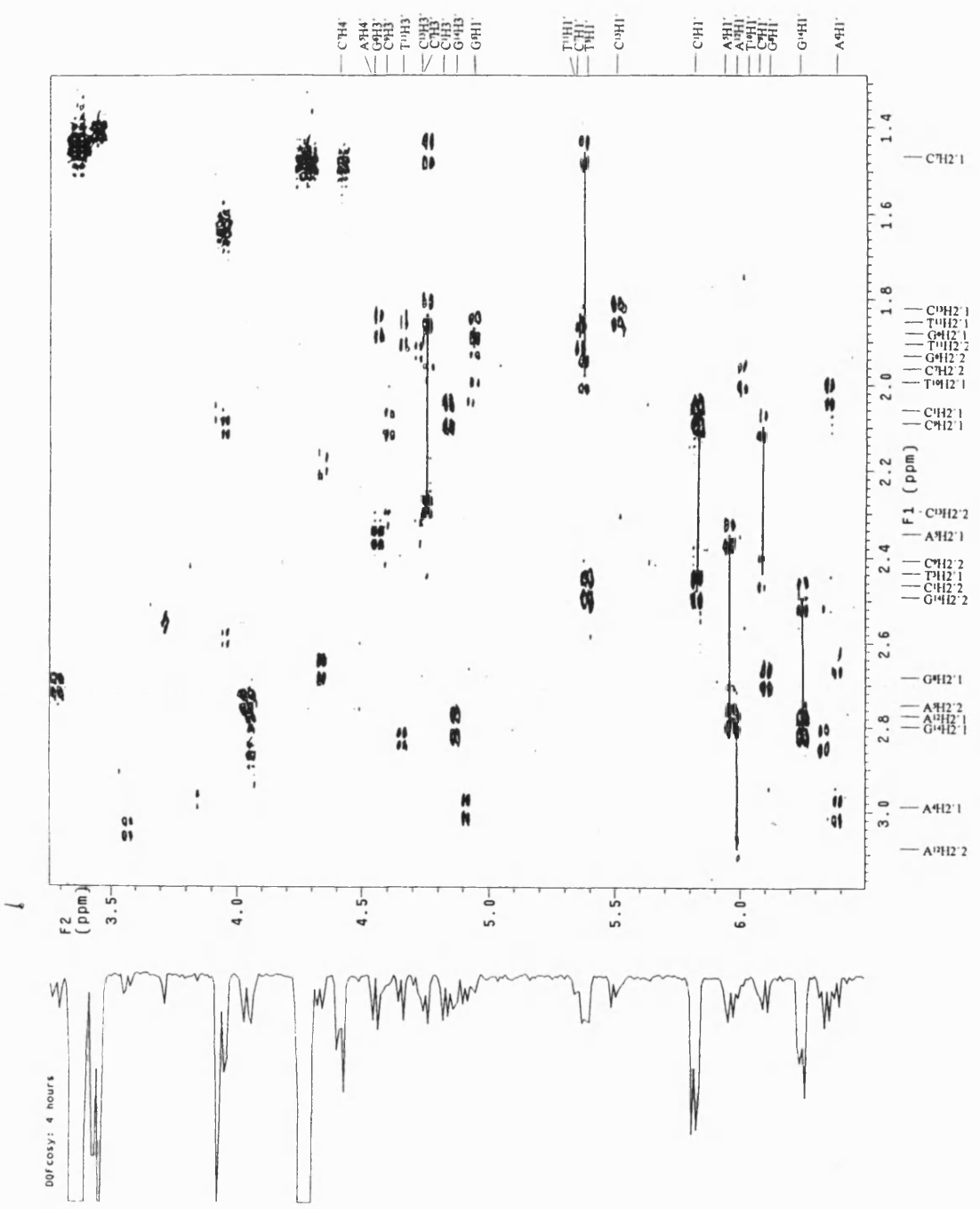


Figure 127 The 600 MHz DQF-COSY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The H1′/H3′→H2′1/H2′2 region.

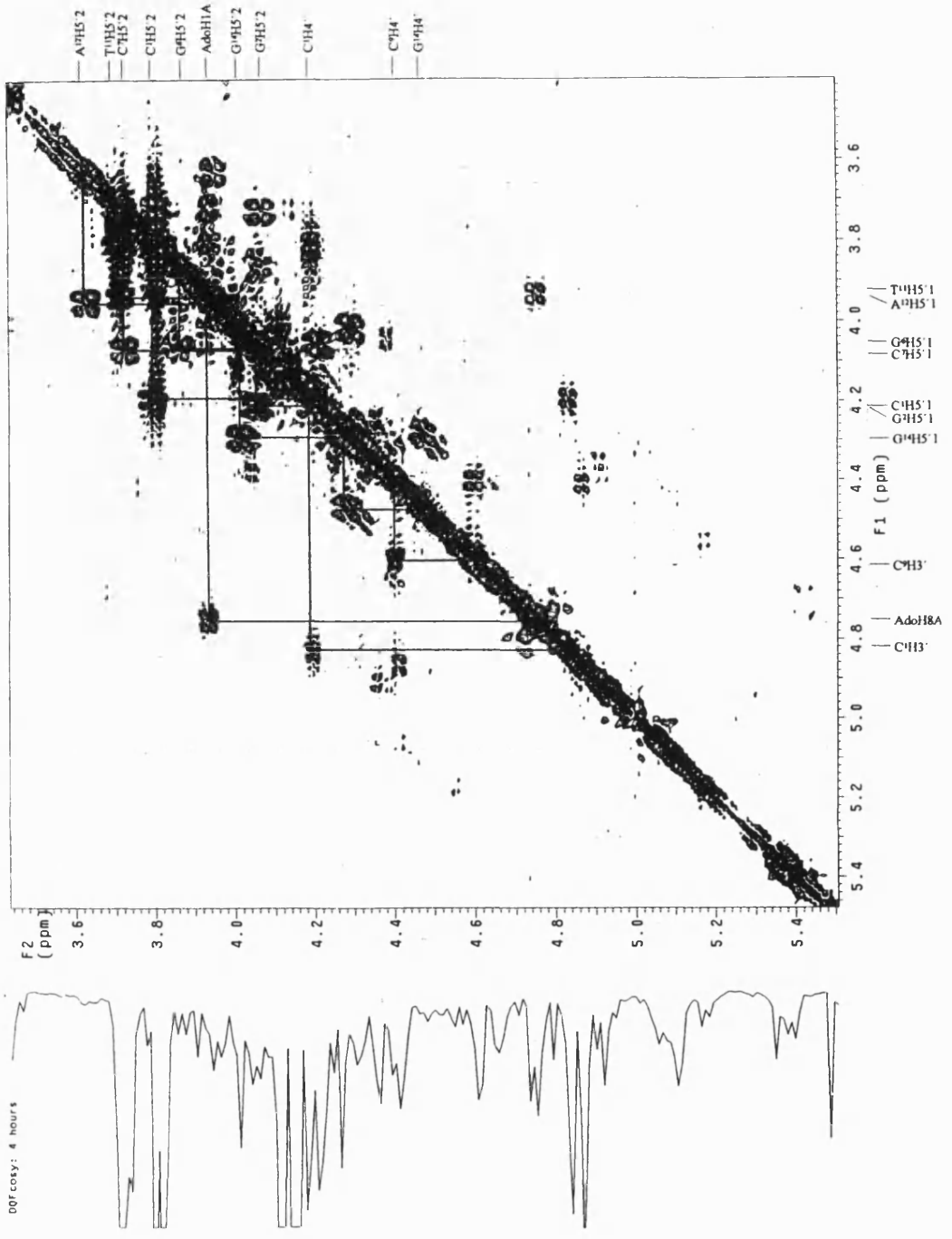


Figure 128 The 600 MHz DQF-COSY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The H5′1→H5′2 region.

Appendix VIII

The 600 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct.

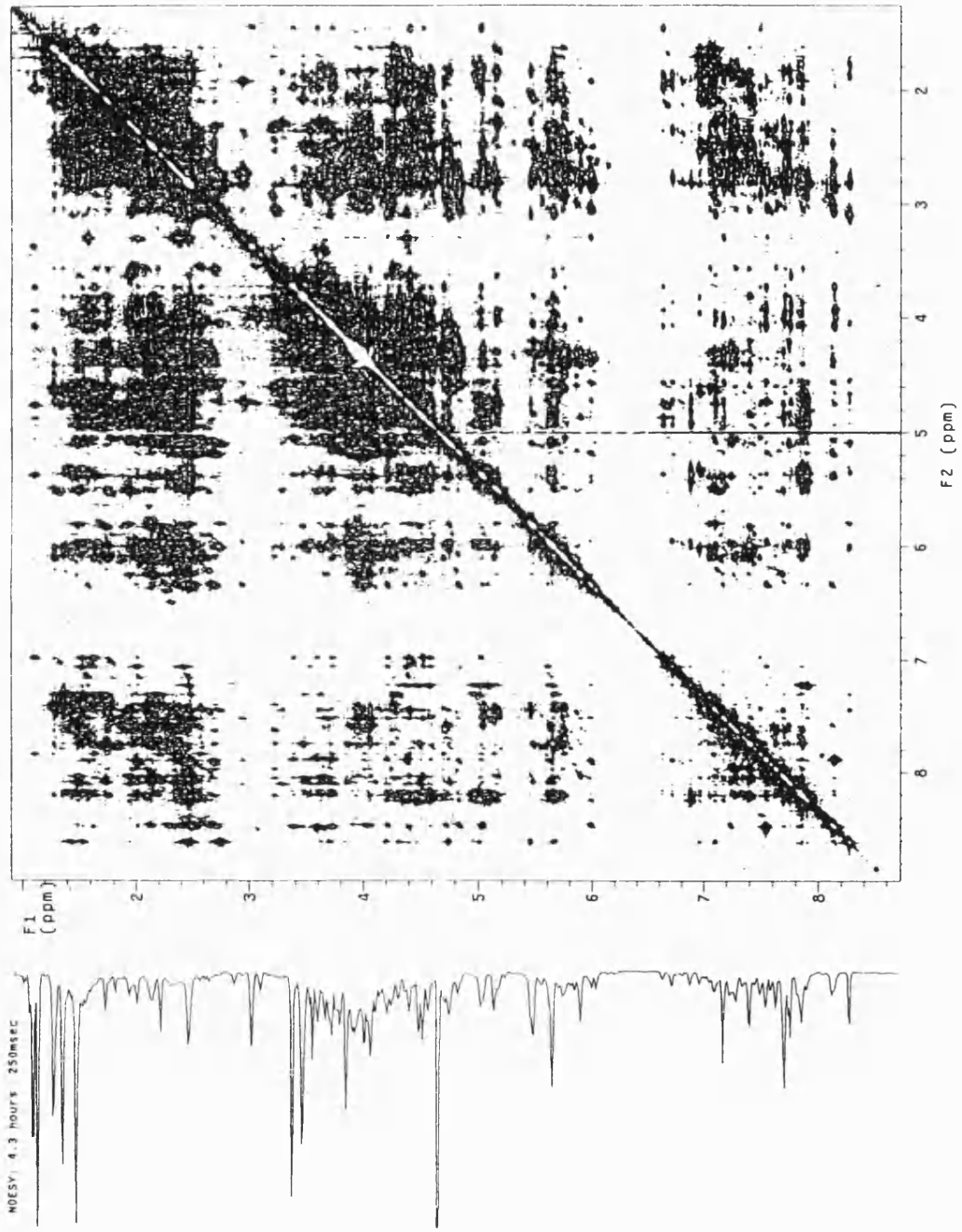


Figure 129 The 600 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct.

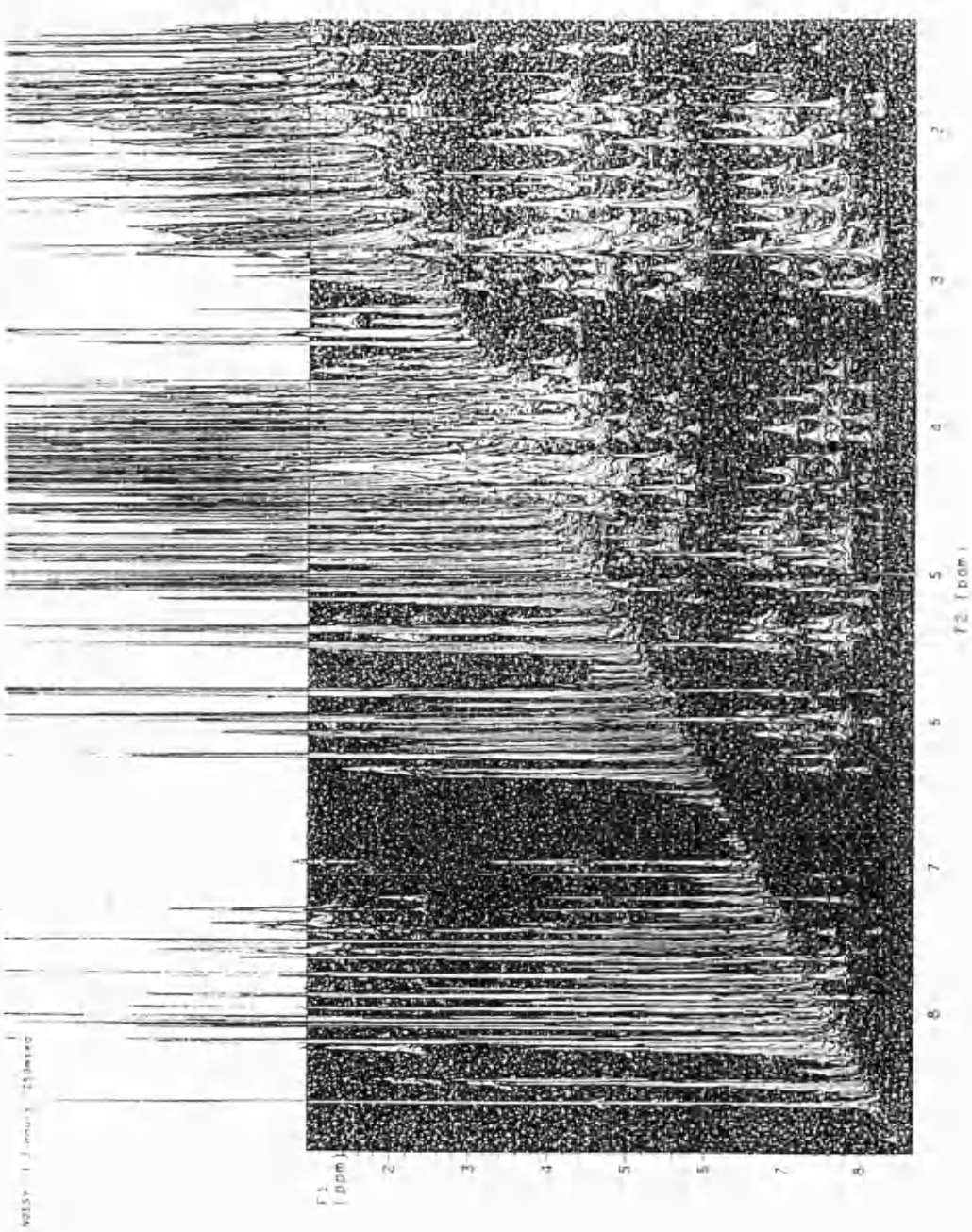


Figure 130 The 600 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14})_2$ DNA adduct. Stack plot.

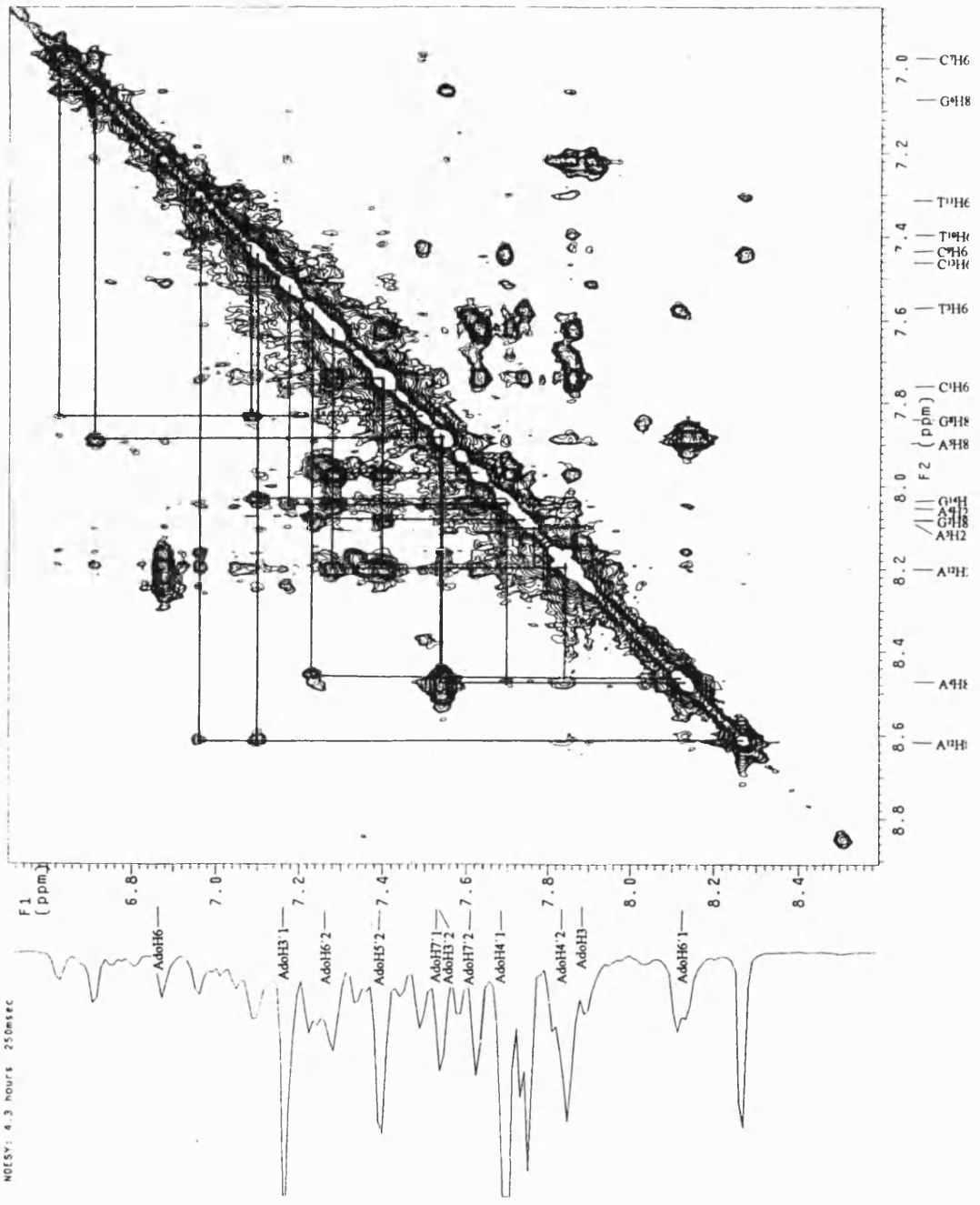


Figure 131 The 600 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14})_{2}$ DNA adduct. The aromatics→aromatics region.

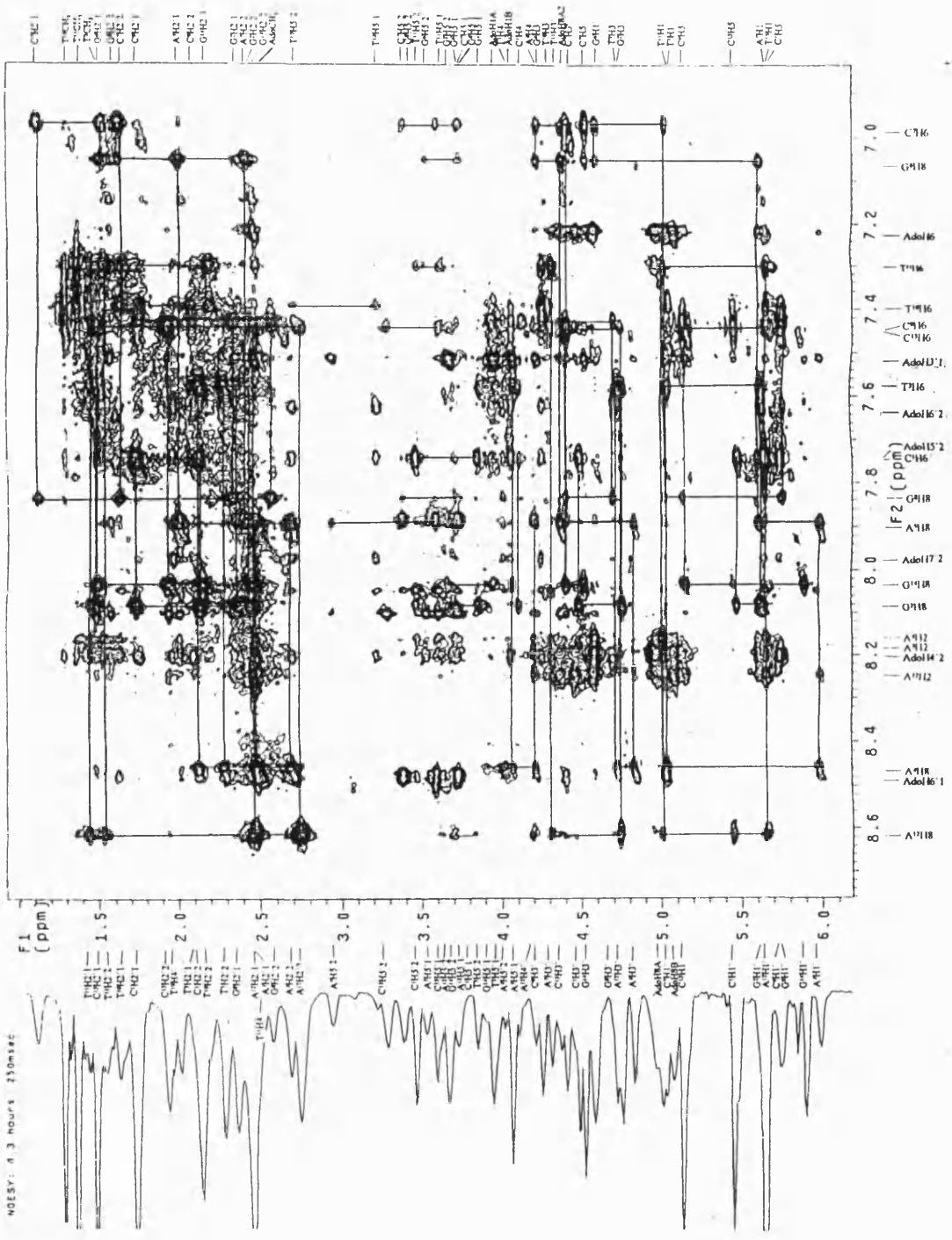


Figure 132 The 600 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5'-
$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14})_2$ DNA adduct. The aromatics→deoxyribose proton NOEs.

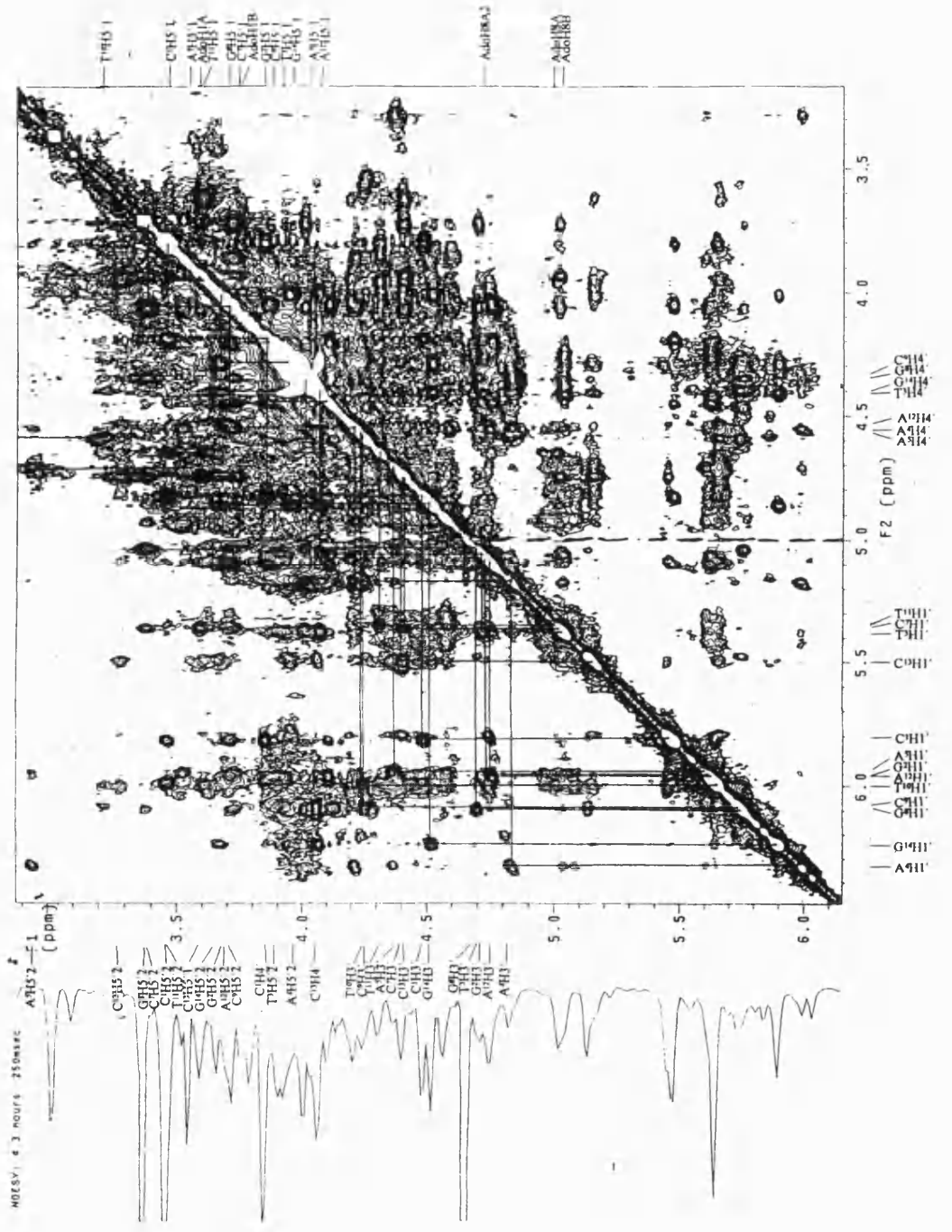


Figure 133 The 600 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-
$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14})_{2}$ DNA adduct. The H1′, H4′, H3′, H5′1 and H5′2 proton NOEs.

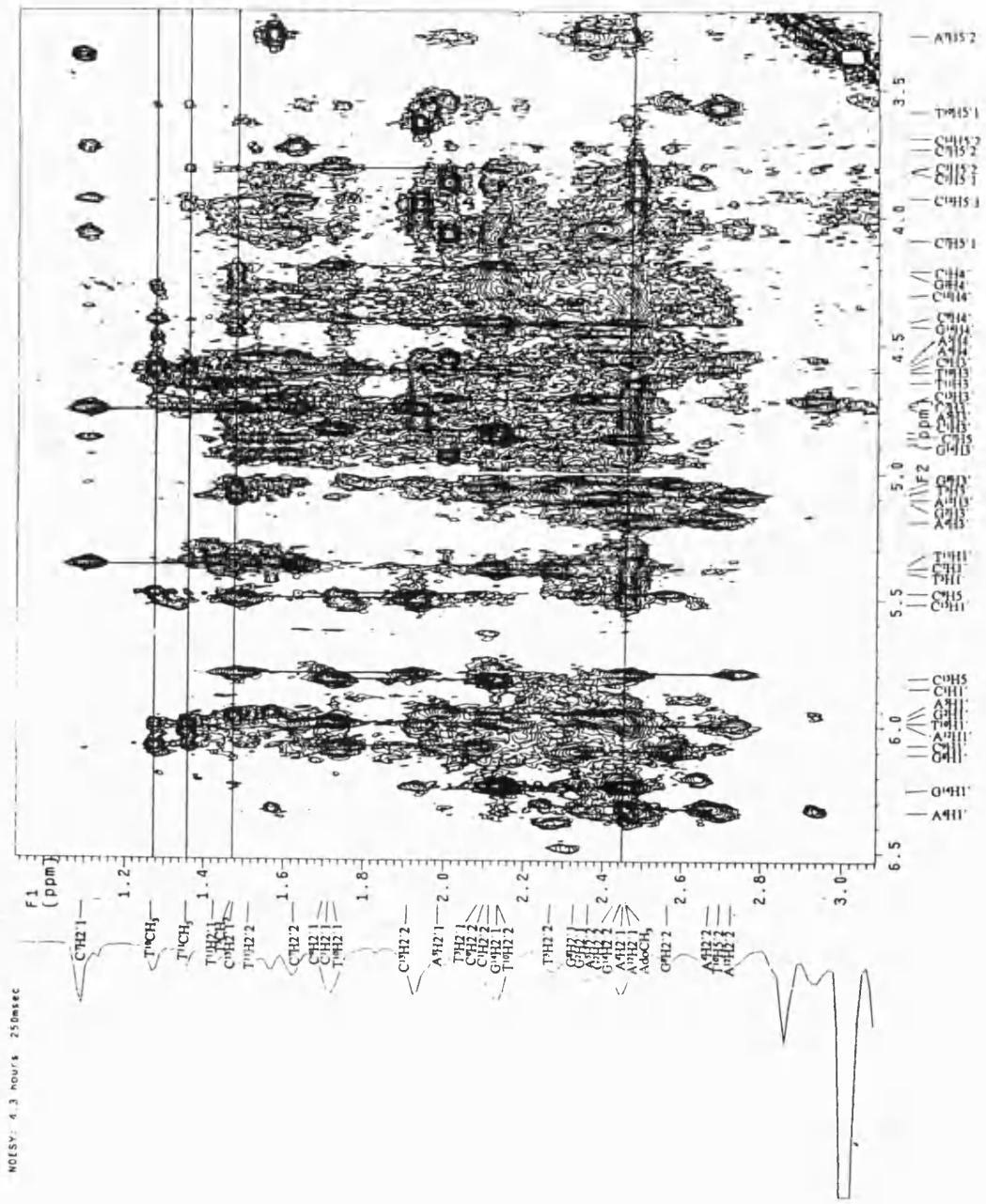


Figure 134 The 600 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The H1′, H3′, H4′→H2′1/H2′ region.

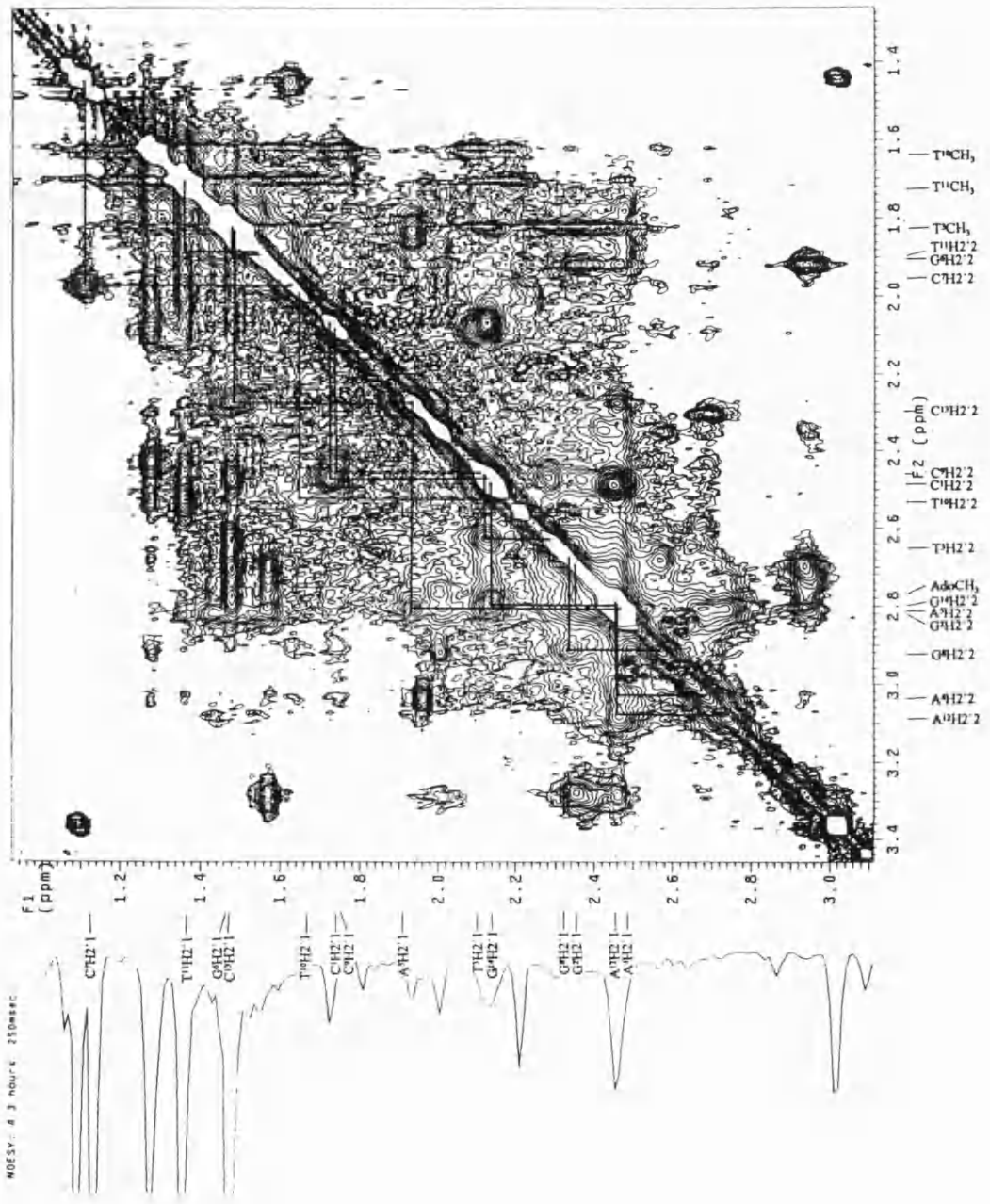


Figure 135 The 600 MHz NOESY ^{1}H NMR spectrum of the adozelesin-5'-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14})_2$ DNA adduct. The H2′1→H2′2 region.

APPENDIX IX

The 600 MHz TOCSY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct.

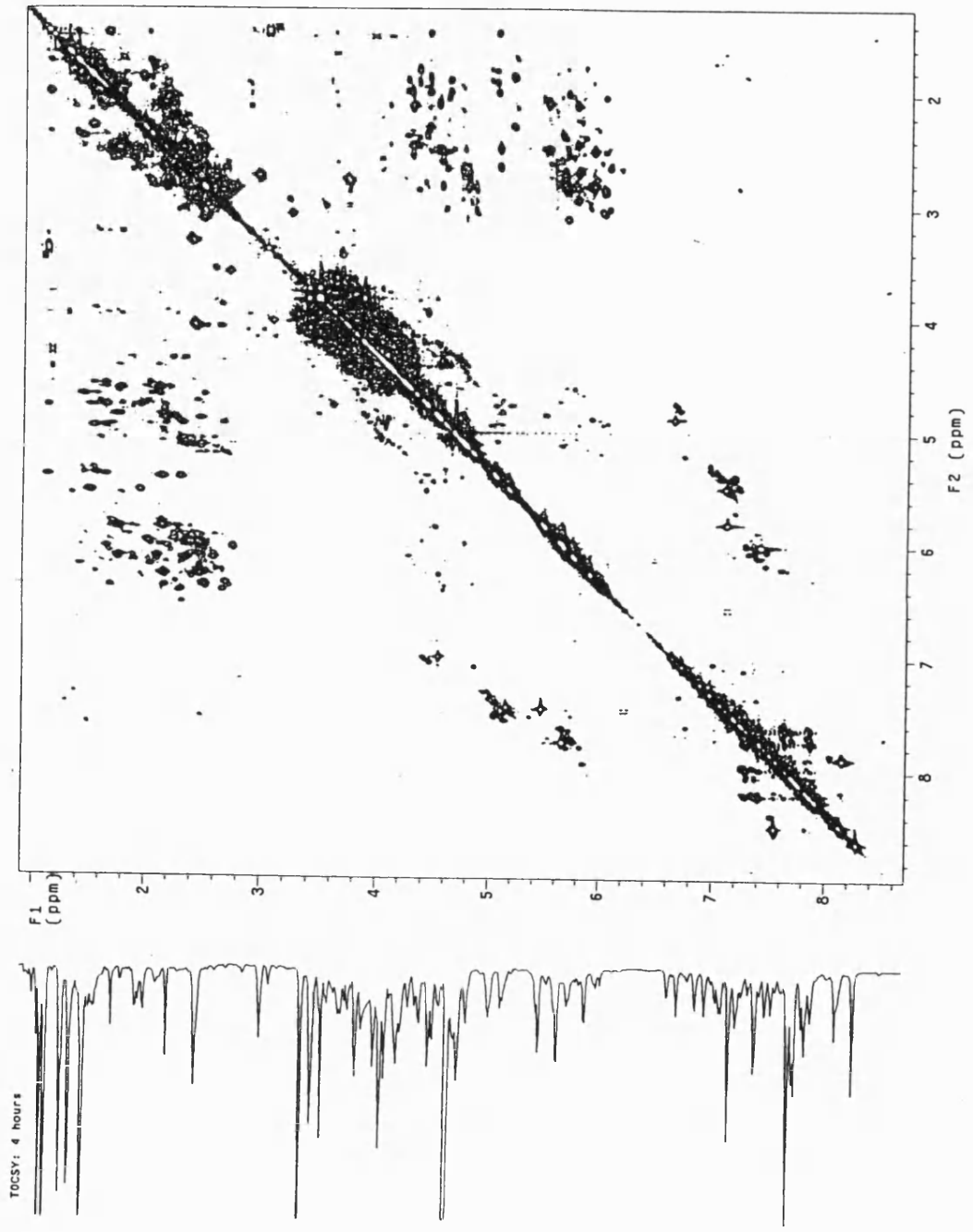


Figure 136 The 600 MHz TOCSY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct.

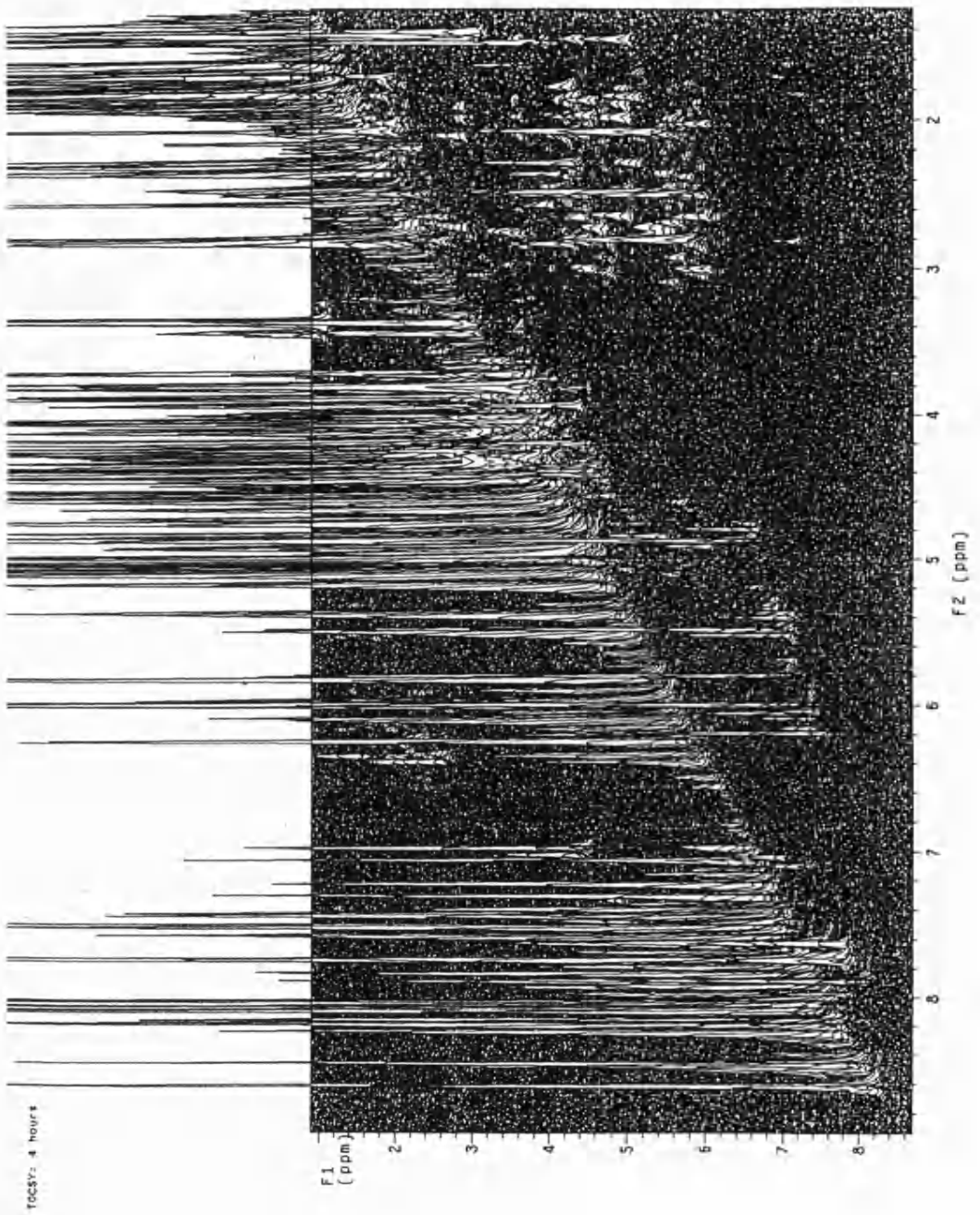


Figure 137 The 600 MHz TOCSY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. Stacked plot.

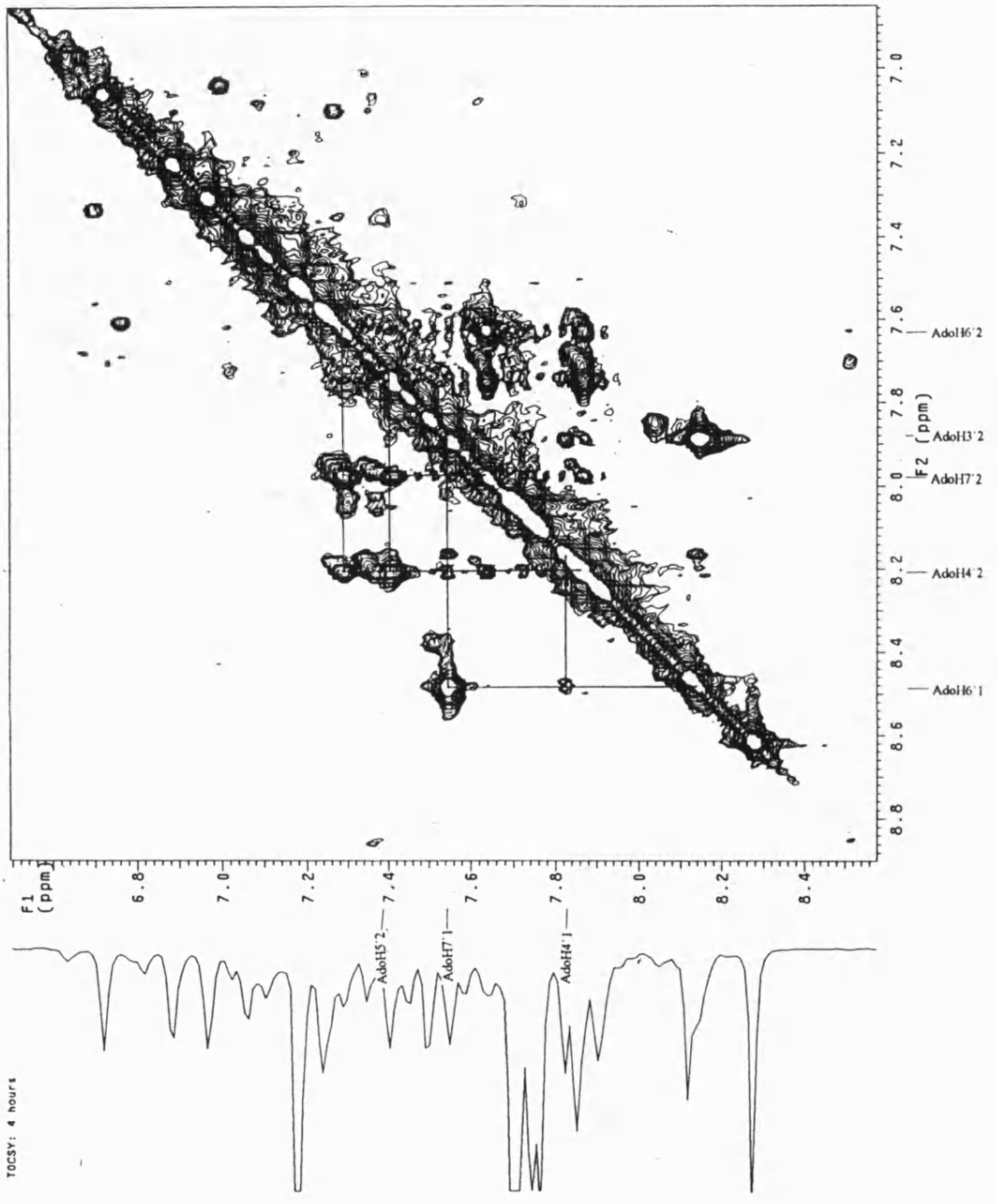


Figure 138 The 600 MHz TOCSY ^{1}H NMR spectrum of the adozelesin-5'-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The aromatics→aromatics region.

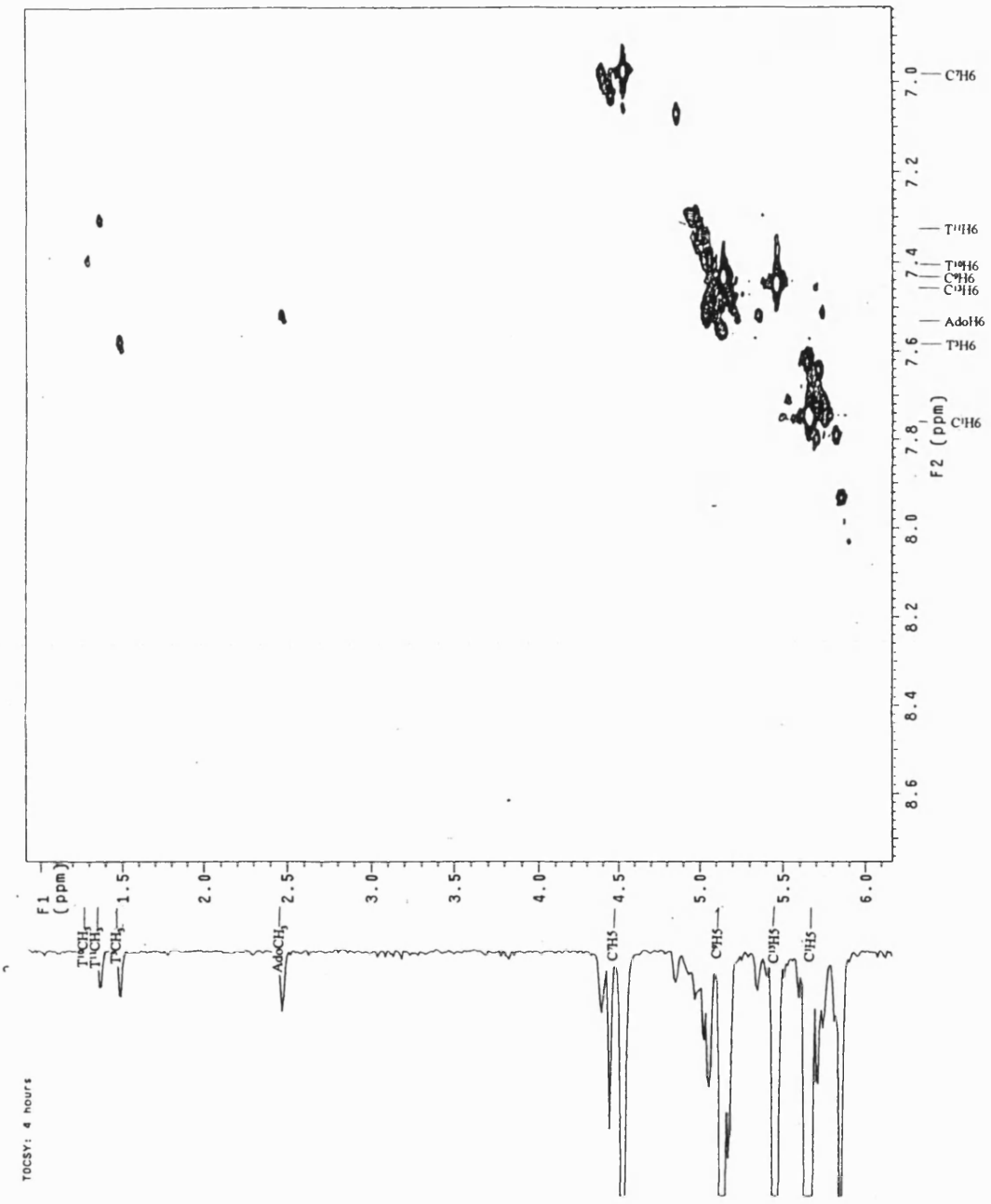


Figure 139 The 600 MHz TOCSY ^{1}H NMR spectrum of the adozelesin-5′-
$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14})_2$ DNA adduct. The aromatics→CH5/methyl region.

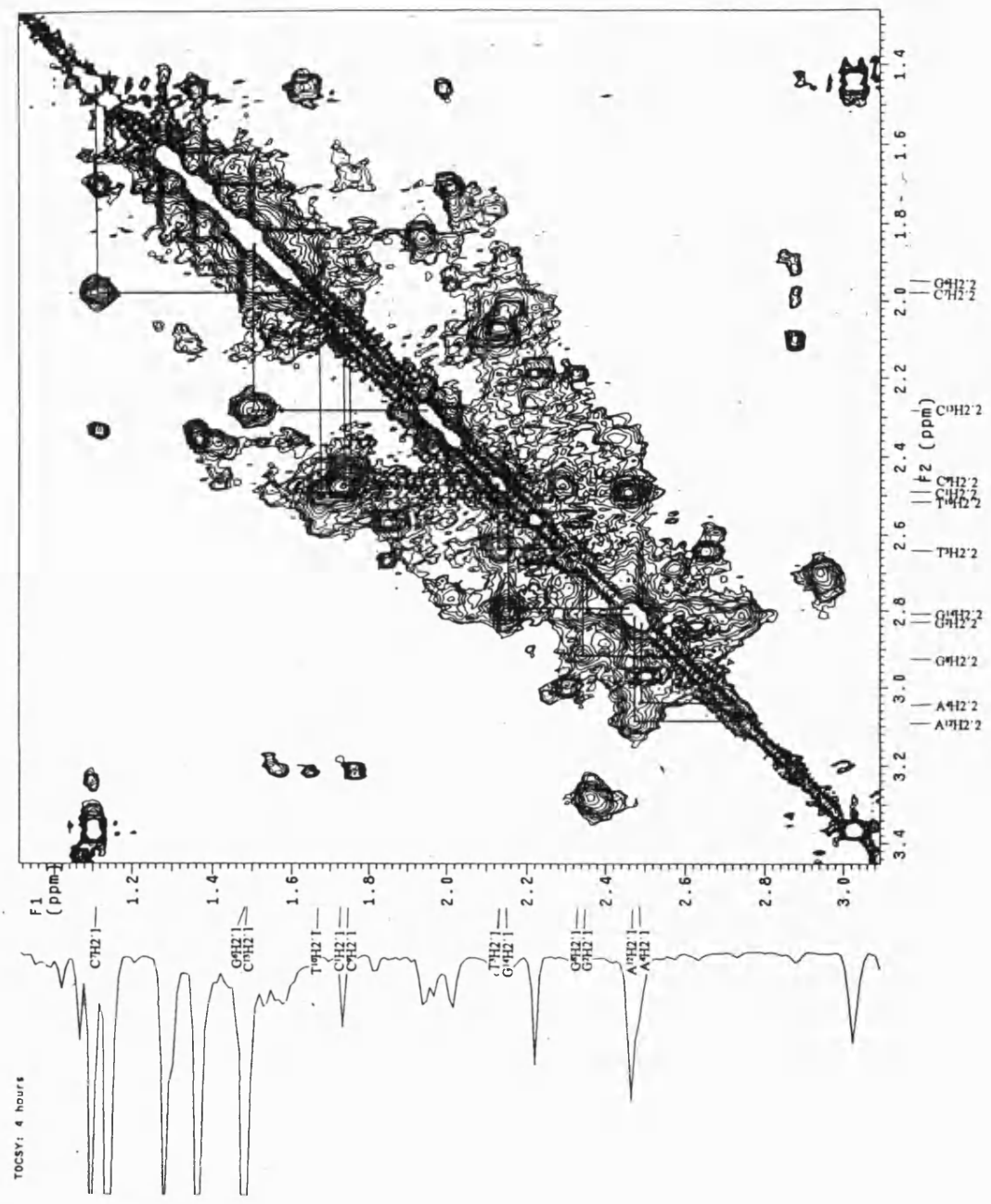


Figure 140 The 600 MHz TOCSY ^{1}H NMR spectrum of the adozelesin-5'-($C^1G^2T^3A^4A^5G^6C^7G^8C^9T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The H2′1→H2′2 region.

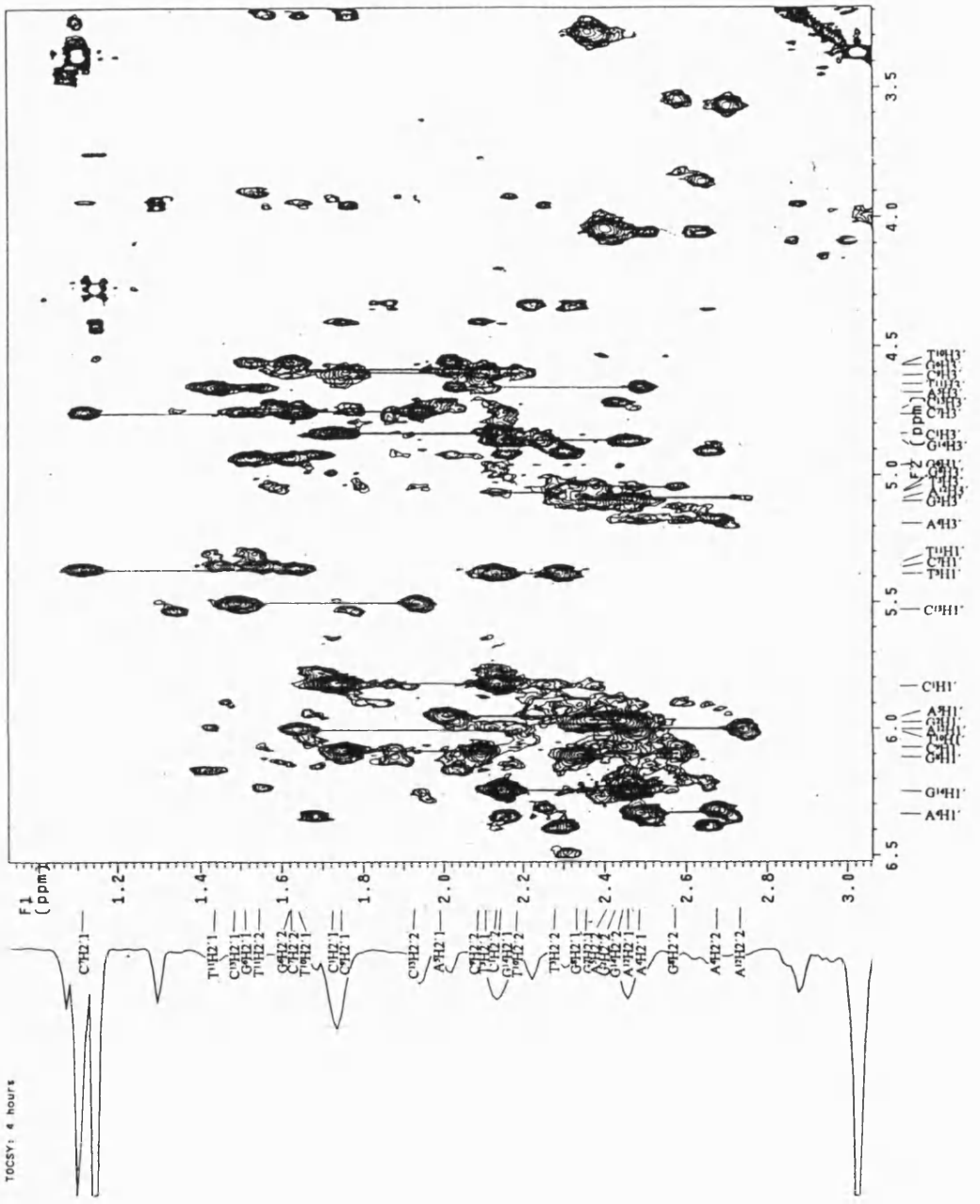


Figure 141 The 600 MHz TOCSY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The H1′/H3′→H2′1/H2′2 region.

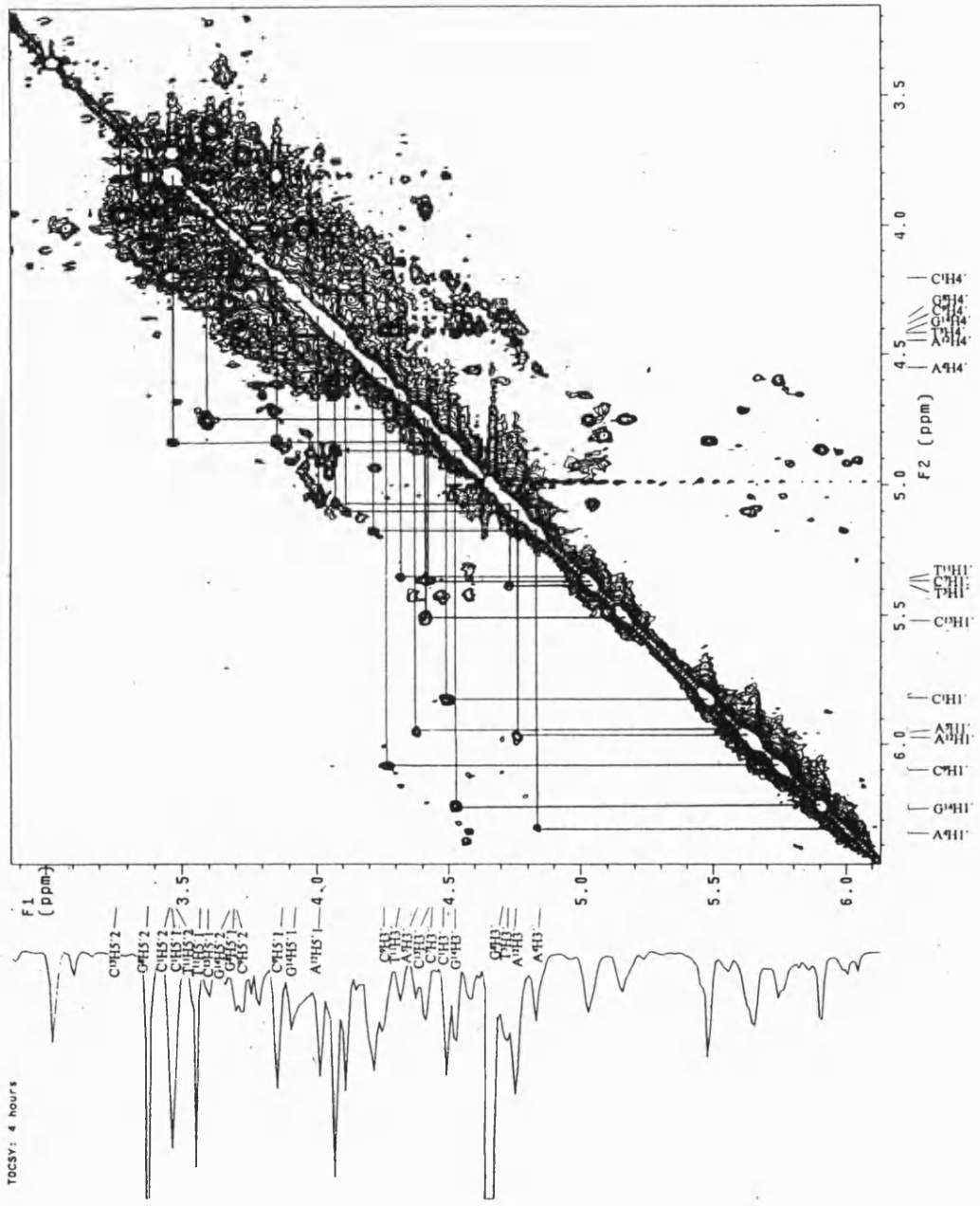


Figure 142 The 600 MHz TOCSY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. Deoxyribose proton NOEs.

Appendix X

The 600 MHz ROESY 2D ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct.

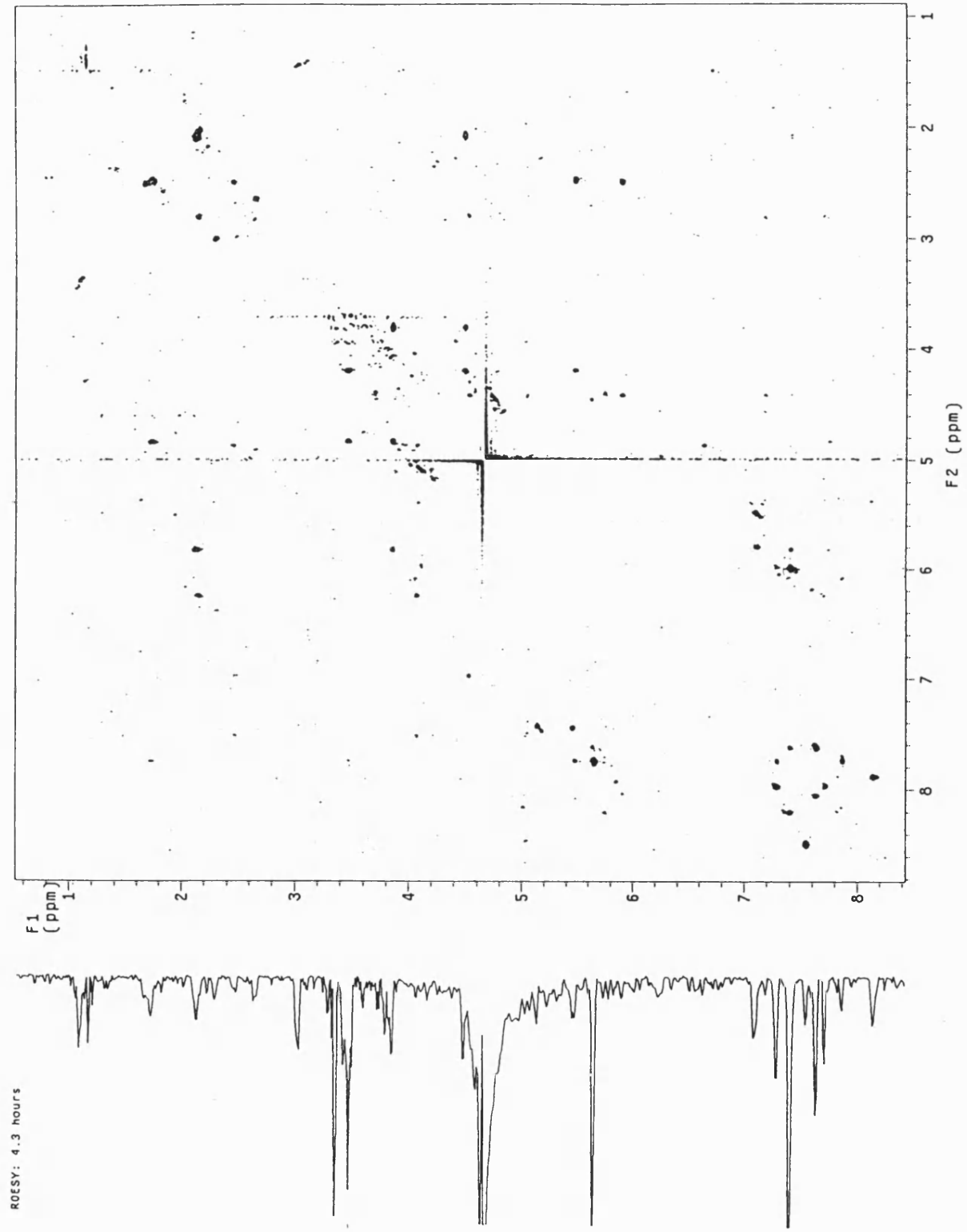


Figure 143 The 600 MHz ROESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct.

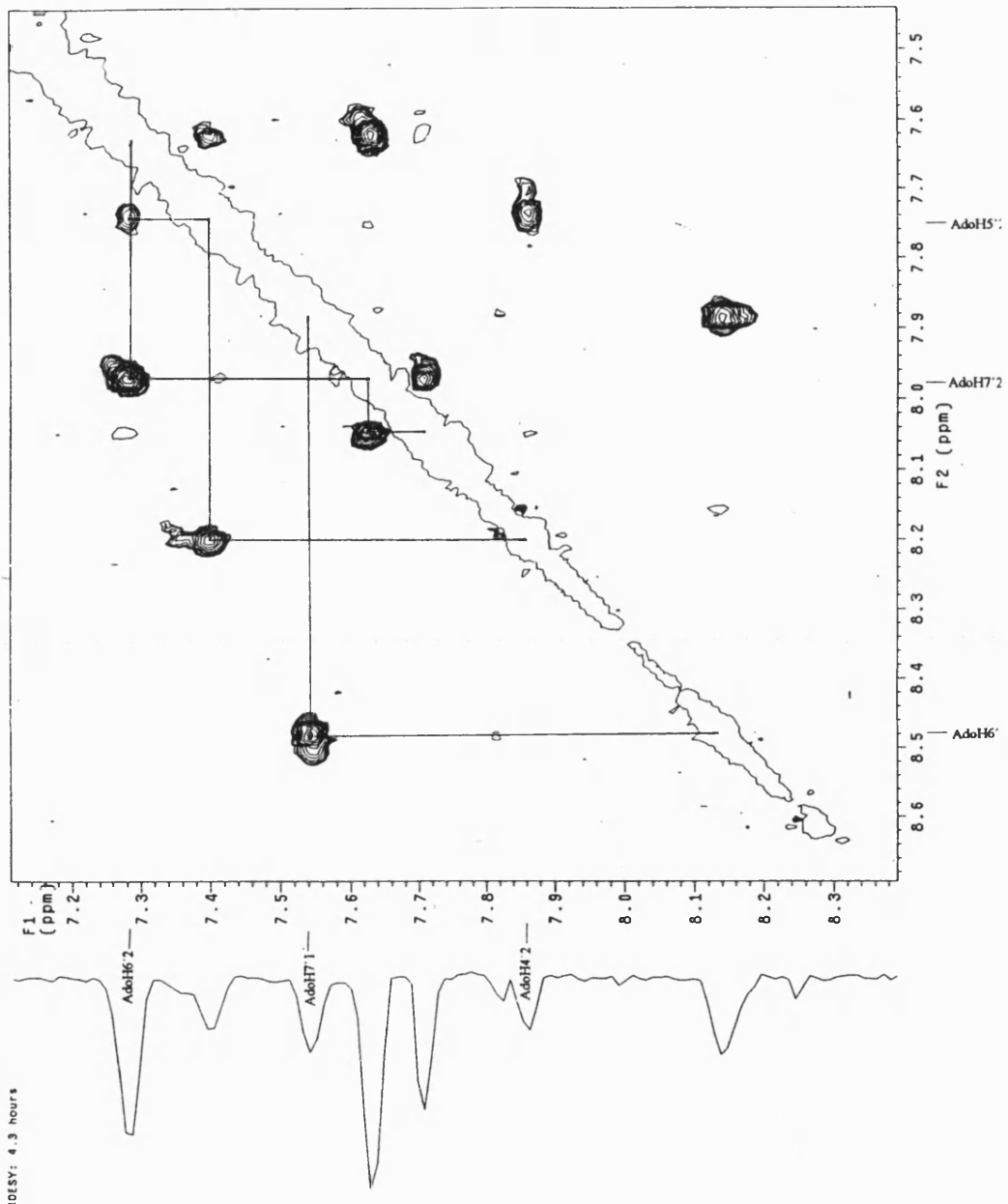


Figure 144 The 600 MHz ROESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The aromatic→aromatic region.

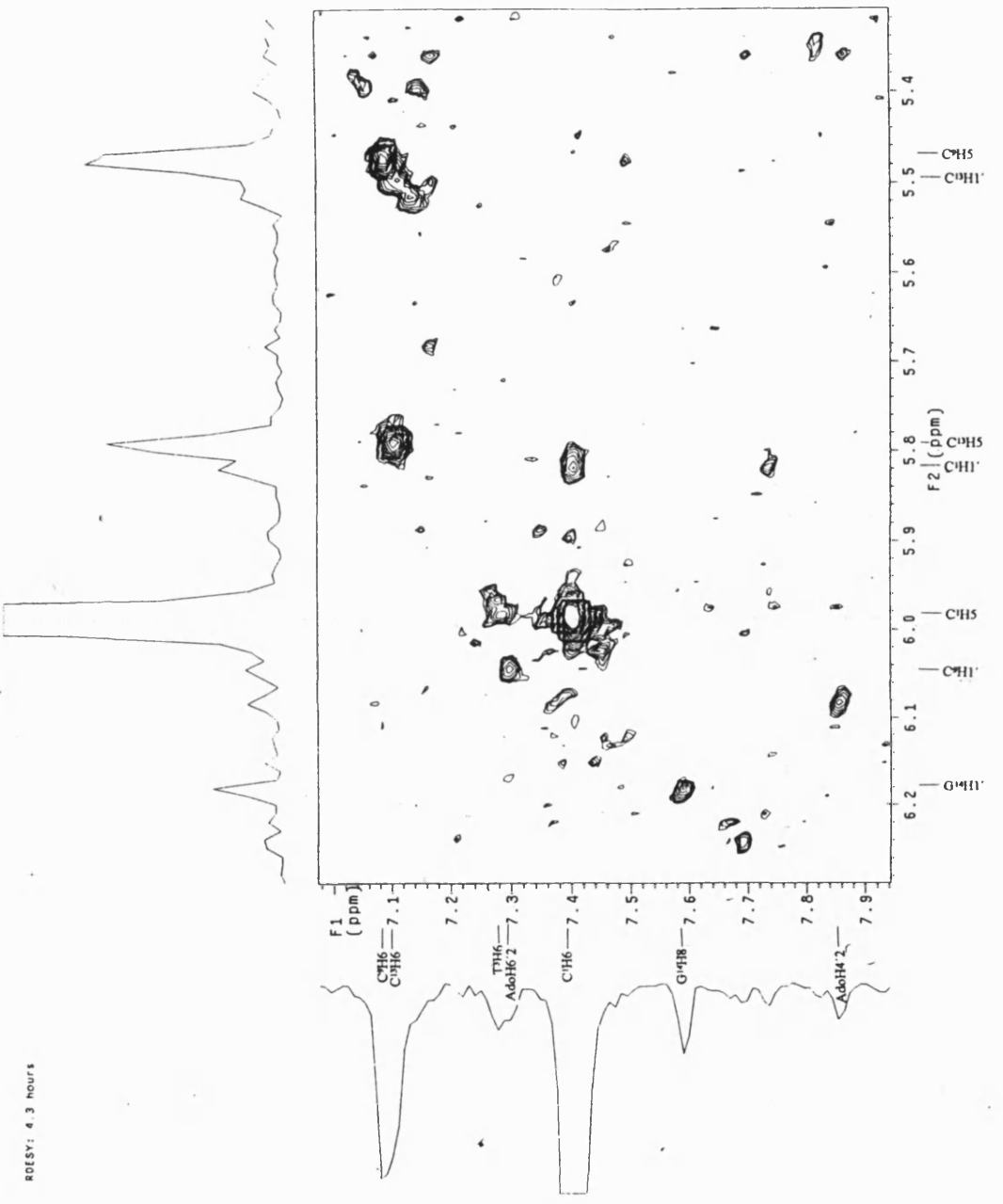


Figure 145 The 600 MHz ROESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14}$)$_2$ DNA adduct. The aromatics→CH5 region.

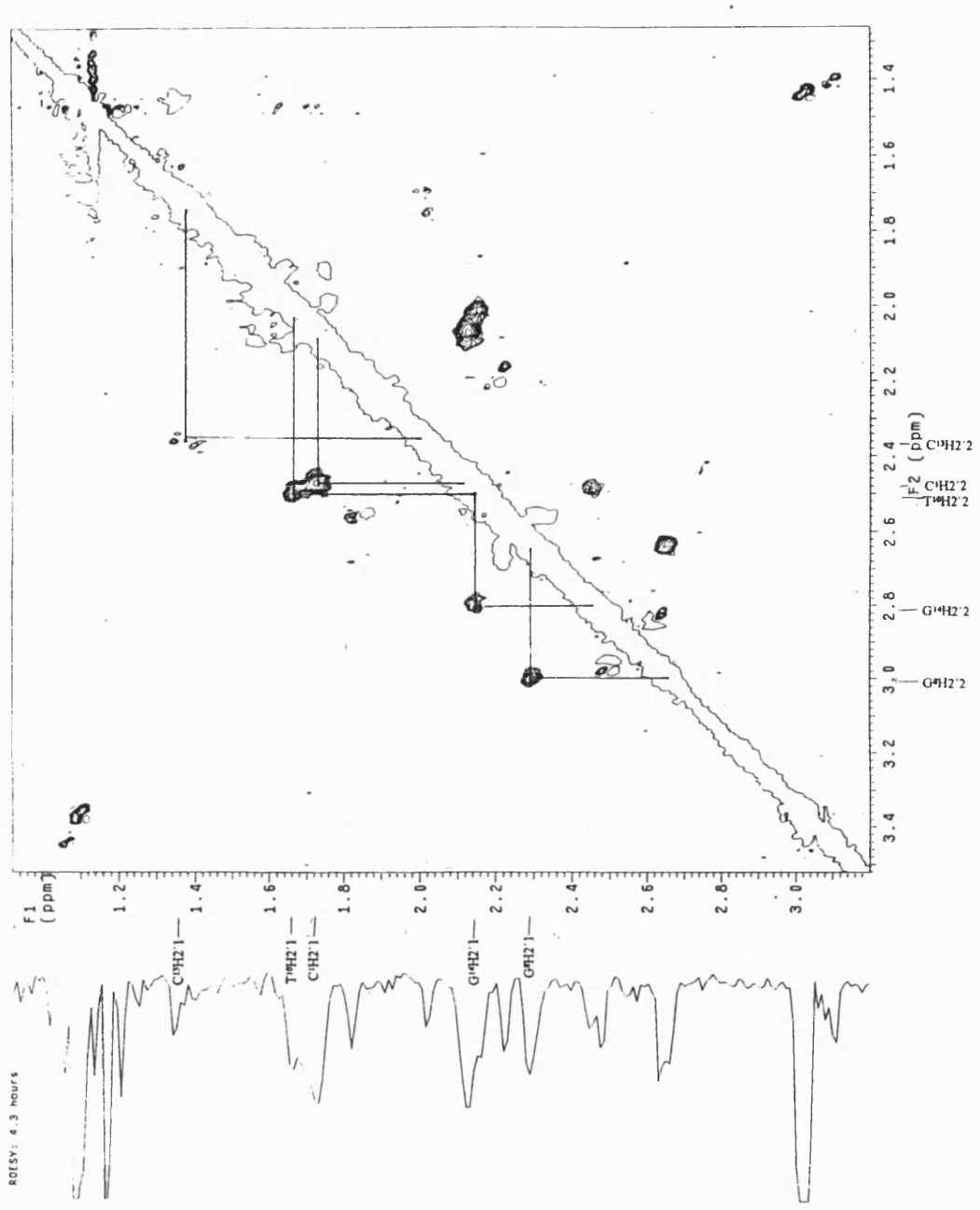


Figure 146 The 600 MHz ROESY ^{1}H NMR spectrum of the adozelesin-5′-$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14})_2$ DNA adduct. The H2′1→H2′2 region.

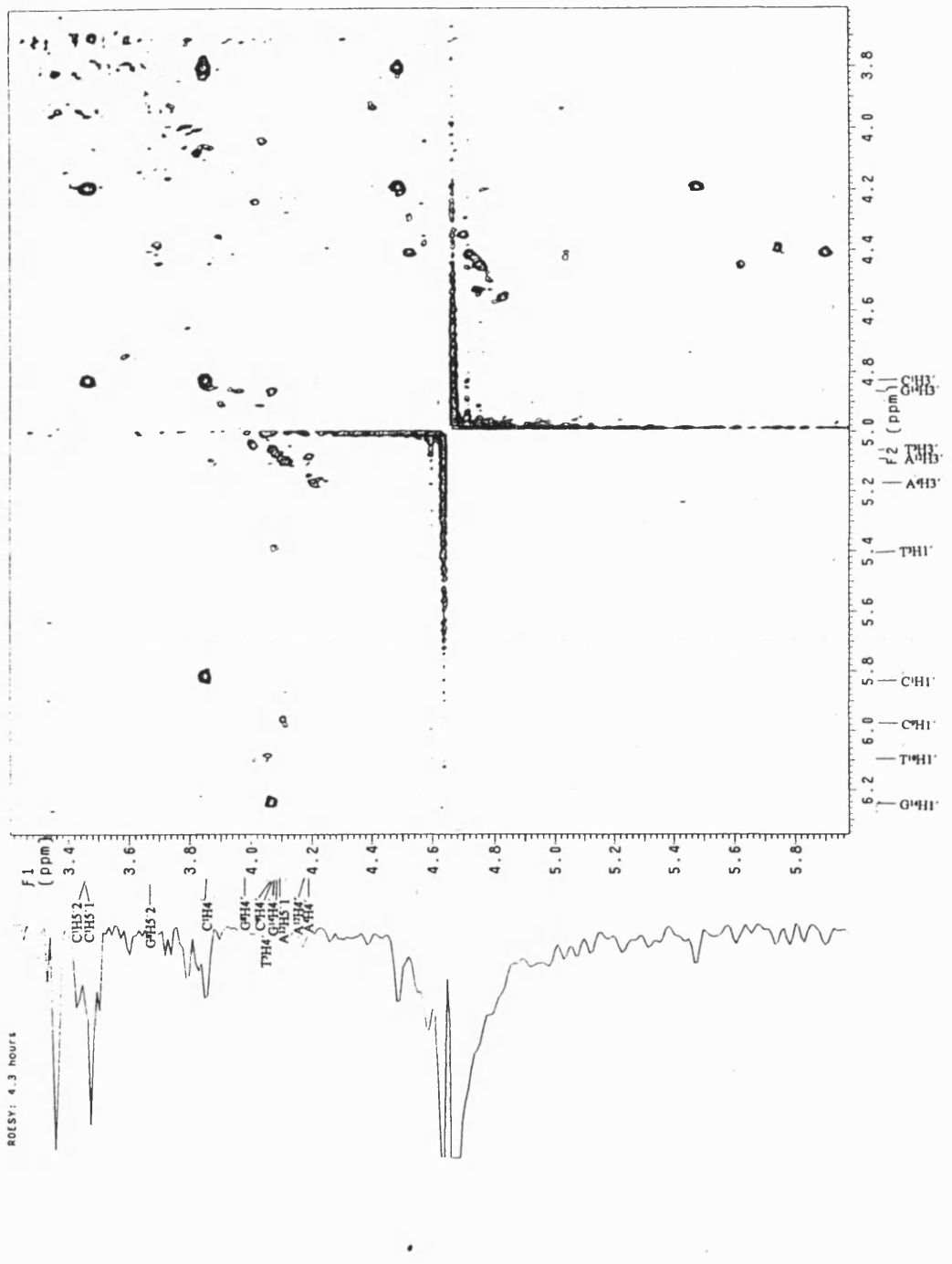


Figure 147 The 600 MHz ROESY ^{1}H NMR spectrum of the adozelesin-5´-
$(C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12*}C^{13}G^{14})_{2}$ DNA adduct. The deoxyribose protons cross-connectivities.

Appendix XI

The Master Assignment Spreadsheet for the adozelesin self-complementary adduct.

Table 6 The Master Assignment spreadsheet.

	1: SHIFT
1: A/C1.H5'2	3.922
2: A/C1.H5'1	3.547
3: A/C1.H4'	3.934
4: A/C1.H1'	5.577
5: A/C1.H2'2	1.803
6: A/C1.H2'1	2.202
7: A/C1.H3'	4.567
8: A/C1.H6	7.512
9: A/C1.H5	5.744
10: A/G2.H5'2	3.951
11: A/G2.H5'1	3.790
12: A/G2.H4'	3.970
13: A/G2.H1'	5.716
14: A/G2.H2'2	2.457
15: A/G2.H2'1	2.508
16: A/G2.H3'	4.833
17: A/G2.H8	7.829
18: A/T3.H5'2	4.022
19: A/T3.H5'1	3.975
20: A/T3.H4'	4.149
21: A/T3.H1'	5.135
22: A/T3.H2'2	2.192
23: A/T3.H2'1	2.367
24: A/T3.H3'	4.798
25: A/T3.H6	7.330
26: A/T3.H5M3	1.555
27: A/T3.H5M2	1.555
28: A/T3.H5M1	1.555
29: A/T3.Q5M	?
30: A/A4.H5'2	4.101
31: A/A4.H5'1	4.061
32: A/A4.H4'	4.284
33: A/A4.H1'	6.067
34: A/A4.H2'2	2.574
35: A/A4.H2'1	2.773

	1: SHIFT
36: A/A4.H3'	4.908
37: A/A4.H8	8.223
38: A/A4.H62	?
39: A/A4.H2	7.809
40: A/A5.H5'2	3.664
41: A/A5.H5'1	3.023
42: A/A5.H4'	4.295
43: A/A5.H1'	5.697
44: A/A5.H2'2	2.097
45: A/A5.H2'1	2.484
46: A/A5.H3'	4.449
47: A/A5.H8	7.648
48: A/A5.H62	?
49: A/A5.H61	?
50: A/A5.H2	7.858
51: A/G6.H1'	5.764
52: A/G6.H2'2	1.601
53: A/G6.H2'1	1.681
54: A/G6.H3'	4.463
55: A/G6.H8	6.822
56: A/C7.H5'2	3.796
57: A/C7.H5'1	3.454
58: A/C7.H4'	4.099
59: A/C7.H1'	5.114
60: A/C7.H2'2	1.200
61: A/C7.H2'1	1.718
62: A/C7.H3'	4.497
63: A/C7.H6	6.737
64: A/C7.H5	4.605
65: B/G8.H5'2	3.796
66: B/G8.H5'1	3.447
67: B/G8.H4'	4.084
68: B/G8.H1'	5.832
69: B/G8.H2'2	2.411
70: B/G8.H2'1	2.649

	1: SHIFT
71: B/G8.H3'	4.787
72: B/G8.H8	7.583
73: B/C9.H5'2	4.027
74: B/C9.H5'1	3.808
75: B/C9.H4'	4.144
76: B/C9.H1'	5.816
77: B/C9.H2'2	1.825
78: B/C9.H2'1	2.175
79: B/C9.H3'	4.337
80: B/C9.H6	7.194
81: B/C9.H5	5.221
82: B/T10.H5'2	3.304
83: B/T10.H5'1	2.789
84: B/T10.H4'	4.151
85: B/T10.H1'	5.748
86: B/T10.H2'2	1.725
87: B/T10.H2'1	2.265
88: B/T10.H3'	4.338
89: B/T10.H6	7.157
90: B/T10.H5M3	1.365
91: B/T10.H5M2	1.365
92: B/T10.H5M1	1.365
93: B/T10.Q5M	?
94: B/T11.H5'2	3.708
95: B/T11.H5'1	3.555
96: B/T11.H4'	3.988
97: B/T11.H1'	5.095
98: B/T11.H2'2	1.526
99: B/T11.H2'1	1.623
100: B/T11.H3'	4.392
101: B/T11.H6	7.062
102: B/T11.H5M3	1.453
103: B/T11.H5M2	1.453
104: B/T11.H5M1	1.453
105: B/T11.Q5M	?

	1: SHIFT
106: B/A12.H5'2	4.182
107: B/A12.H5'1	3.788
108: B/A12.H4'	4.271
109: B/A12.H1'	5.749
110: B/A12.H2'2	2.542
111: B/A12.H2'1	2.809
112: B/A12.H3'	4.832
113: B/A12.H8	8.375
114: B/A12.H2	7.956
115: B/C13.H5'2	3.671
116: B/C13.H5'1	3.353
117: B/C13.H4'	4.023
118: B/C13.H1'	5.240
119: B/C13.H2'2	1.575
120: B/C13.H2'1	2.015
121: B/C13.H3'	4.493
122: B/C13.H6	7.212
123: B/C13.H5	5.538
124: B/G14.H5'2	4.033
125: B/G14.H5'1	3.759
126: B/G14.H4'	4.145
127: B/G14.H1'	5.993
128: B/G14.H2'2	2.226
129: B/G14.H2'1	2.527
130: B/G14.H3'	4.596
131: B/G14.H8	7.797
132: A/OH1.H3	7.985
133: A/OH1.H6	6.980
134: A/OH1.M1	2.546
135: A/OH1.M2	2.546
136: A/OH1.M3	2.546
137: A/OH1.H8A2	4.495
138: A/OH1.H8A	5.063
139: A/OH1.H8B	5.155
140: A/OH1.H1A	3.689

	1: SHIFT
141: A/OH1.H1B	3.812
142: A/OH1.H7'1	7.629
143: A/OH1.H6'1	8.243
144: A/OH1.H4'1	7.805
145: A/OH1.H3'1	7.283
146: A/OH1.H4'2	7.932
147: A/OH1.H5'2	7.497
148: A/OH1.H6'2	7.373
149: A/OH1.H7'2	7.736
150: A/OH1.H3'2	7.651

APPENDIX XII

The Peak Pick Spreadsheet for the adozelesin-DNA self-complementary adduct.

Table 7 The Peak Pick Spreadsheet.

	1: PCS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
1: P1	7.51	1.80	4.77e+06	A/C1.H6	A/C1.H2'2	1.58	3.00
2: P2	7.51	2.20	5.83e+06	A/C1.H6	A/C1.H2'1	1.58	3.00
3: P3	7.83	2.51	6.77e+06	A/G2.H8	A/G2.H2'1	1.58	3.00
4: P4	7.83	2.46	2.37e+06	A/G2.H8	A/G2.H2'2	1.58	3.00
5: P5	7.33	2.19	5.68e+06	A/T3.H6	A/T3.H2'2	1.58	3.00
6: P6	7.33	2.37	6.29e+06	A/T3.H6	A/T3.H2'1	1.58	3.00
7: P7	8.23	2.58	4.81e+06	A/A4.H8	A/A4.H2'2	1.58	3.00
8: P8	8.23	2.77	2.69e+06	A/A4.H8	A/A4.H2'1	1.58	3.00
9: P9	7.65	2.09	3.5e+06	A/A5.H8	A/A5.H2'2	1.58	3.00
10: P10	7.65	2.48	2.94e+06	A/A5.H8	A/A5.H2'1	1.58	3.00
11: P11	6.82	1.60	6.62e+05	A/G6.H8	A/G6.H2'2	1.58	5.00
12: P12	6.82	1.68	4.2e+05	A/G6.H8	A/G6.H2'1	1.58	5.00
13: P13	6.73	1.20	8.41e+05	A/C7.H6	A/C7.H2'2	1.58	5.00
14: P14	6.73	1.72	7.44e+05	A/C7.H6	A/C7.H2'1	1.58	5.00
15: P15	7.59	2.41	2.46e+06	B/G8.H8	B/G8.H2'2	1.58	3.00
16: P16	7.59	2.65	1.19e+06	B/G8.H8	B/G8.H2'1	1.58	3.00
17: P17	7.19	2.17	2.13e+06	B/C9.H6	B/C9.H2'1	1.58	3.00
18: P18	7.19	1.82	3.9e+06	B/C9.H6	B/C9.H2'2	1.58	3.00
19: P19	7.15	2.26	2.79e+06	B/T10.H6	B/T10.H2'1	1.58	3.00
20: P20	7.15	1.72	4.1e+06	B/T10.H6	B/T10.H2'2	1.58	3.00
21: P21	7.06	1.52	1.13e+06	B/T11.H6	B/T11.H2'2	1.58	3.00
22: P22	7.06	1.62	1.54e+06	B/T11.H6	B/T11.H2'1	1.58	3.00
23: P23	8.37	2.54	3.73e+06	B/A12.H8	B/A12.H2'2	1.58	3.00
24: P24	8.37	2.81	3.99e+06	B/A12.H8	B/A12.H2'1	1.58	3.00
25: P25	7.21	1.57	5.99e+06	B/C13.H6	B/C13.H2'2	1.58	3.00
26: P26	7.21	2.01	5.41e+06	B/C13.H6	B/C13.H2'1	1.58	3.00
27: P27	7.79	2.22	3.04e+06	B/G14.H8	B/G14.H2'2	1.58	3.00
28: P28	7.79	2.53	3.9e+06	B/G14.H8	B/G14.H2'1	1.58	3.00
29: P29	7.83	1.80	2.36e+06	A/G2.H8	A/C1.H2'2	1.58	3.00
30: P30	7.83	2.20	3.67e+06	A/G2.H8	A/C1.H2'1	1.58	3.00
31: P31	7.33	2.46	1.07e+06	A/T3.H6	A/G2.H2'2	1.58	3.00
32: P32	7.33	2.51	1.21e+06	A/T3.H6	A/G2.H2'1	1.58	3.00
33: P33	8.22	2.19	7.76e+05	A/A4.H8	A/T3.H2'2	1.58	5.00
34: P34	8.23	2.37	2.5e+06	A/A4.H8	A/T3.H2'1	1.58	3.00
35: P35	7.65	2.77	2.77e+06	A/A5.H8	A/A4.H2'1	1.58	3.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
36: P36	7.65	2.58	1.8e+06	A/A5.H8	A/A4.H2'2	1.58	3.00
37: P37	6.82	2.09	1.11e+06	A/G6.H8	A/A5.H2'2	1.58	3.00
38: P38	6.82	2.49	1.33e+06	A/G6.H8	A/A5.H2'1	1.58	3.00
39: P39	6.74	1.60	8.63e+05	A/C7.H6	A/G6.H2'2	1.58	5.00
40: P40	6.74	1.68	2.43e+05	A/C7.H6	A/G6.H2'1	1.58	5.00
41: P43	7.19	2.65	1.84e+06	B/C9.H6	B/G8.H2'1	1.58	3.00
42: P44	7.19	2.42	1.74e+06	B/C9.H6	B/G8.H2'2	1.58	3.00
43: P45	7.15	2.17	9.36e+05	B/T10.H6	B/C9.H2'1	1.58	5.00
44: P46	7.16	1.82	1.78e+06	B/T10.H6	B/C9.H2'2	1.58	3.00
45: P47	7.06	2.26	3.61e+06	B/T11.H6	B/T10.H2'1	1.58	3.00
46: P48	7.06	1.73	2.44e+06	B/T11.H6	B/T10.H2'2	1.58	3.00
47: P49	8.37	1.53	8.52e+05	B/A12.H8	B/T11.H2'2	1.58	5.00
48: P50	8.37	1.62	5.14e+05	B/A12.H8	B/T11.H2'1	1.58	5.00
49: P51	7.21	2.81	8.28e+05	B/C13.H6	B/A12.H2'1	1.58	5.00
50: P52	7.21	2.54	4.06e+06	B/C13.H6	B/A12.H2'2	1.58	3.00
51: P53	7.79	1.58	2.97e+06	B/G14.H8	B/C13.H2'2	1.58	3.00
52: P54	7.79	2.01	2.16e+06	B/G14.H8	B/C13.H2'1	1.58	3.00
53: P55	7.51	5.57	2.4e+06	A/C1.H6	A/C1.H1'	1.58	3.00
54: P56	7.83	5.57	6.81e+05	A/G2.H8	A/C1.H1'	1.58	5.00
55: P57	7.83	5.71	7.01e+05	A/G2.H8	A/G2.H1'	1.58	5.00
56: P58	7.33	5.72	1.97e+06	A/T3.H6	A/G2.H1'	1.58	3.00
57: P59	7.33	5.13	2.32e+06	A/T3.H6	A/T3.H1'	1.58	3.00
58: P60	8.22	5.14	2.12e+06	A/A4.H8	A/T3.H1'	1.58	3.00
59: P61	8.22	6.07	1.28e+06	A/A4.H8	A/A4.H1'	1.58	3.00
60: P62	7.65	6.07	1.49e+06	A/A5.H8	A/A4.H1'	1.58	3.00
61: P63	7.65	5.70	1.01e+06	A/A5.H8	A/A5.H1'	1.58	3.00
62: P64	6.82	5.70	4.85e+05	A/G6.H8	A/A5.H1'	1.58	5.00
63: P65	6.73	5.11	3.95e+05	A/C7.H6	A/C7.H1'	1.58	5.00
64: P66	7.58	5.83	8.92e+05	B/G8.H8	B/G8.H1'	1.58	5.00
65: P67	7.20	5.84	1.23e+06	B/C9.H6	B/G8.H1'	1.58	3.00
66: P68	7.19	5.82	1.43e+06	B/C9.H6	B/C9.H1'	1.58	3.00
67: P69	7.16	5.82	1.35e+06	B/T10.H6	B/C9.H1'	1.58	3.00
68: P71	7.15	5.75	1.57e+06	B/T10.H6	B/T10.H1'	1.58	3.00
69: P72	7.06	5.75	1.88e+06	B/T11.H6	B/T10.H1'	1.58	3.00
70: P73	7.06	5.09	3.17e+05	B/T11.H6	B/T11.H1'	1.58	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
71: P74	8.37	5.75	4.78e+05	B/A12.H8	B/A12.H1'	1.58	5.00
72: P75	7.20	5.75	1.22e+06	B/C13.H6	B/A12.H1'	1.58	3.00
73: P76	7.20	5.24	3.19e+06	B/C13.H6	B/C13.H1'	1.58	3.00
74: P77	7.79	5.24	8.78e+05	B/G14.H8	B/C13.H1'	1.58	5.00
75: P78	7.79	5.99	1.79e+06	B/G14.H8	B/G14.H1'	1.58	3.00
76: P79	7.51	7.82	2.56e+05	A/C1.H6	A/G2.H8	1.58	5.00
77: P80	7.83	7.31	7.28e+05	A/G2.H8	A/T3.H6	1.58	5.00
78: P81	8.22	7.31	5.55e+05	A/A4.H8	A/T3.H6	1.58	5.00
79: P82	8.22	7.65	2.21e+06	A/A4.H8	A/A5.H8	1.58	3.00
80: P83	7.64	6.81	5.05e+05	A/A5.H8	A/G6.H8	1.58	5.00
81: P84	6.82	6.73	3.53e+05	A/G6.H8	A/C7.H6	1.58	5.00
82: P85	7.19	7.57	8.73e+05	B/C9.H6	B/G8.H8	1.58	5.00
83: P87	8.37	7.05	6.31e+04	B/A12.H8	B/T11.H6	1.58	7.00
84: P88	7.21	8.36	5.66e+05	B/C13.H6	B/A12.H8	1.58	5.00
85: P89	7.79	7.21	6.06e+05	B/G14.H8	B/C13.H6	1.58	5.00
86: P90	7.51	4.57	2.88e+06	A/C1.H6	A/C1.H3'	1.58	3.00
87: P91	7.83	4.57	5.41e+05	A/G2.H8	A/C1.H3'	1.58	5.00
88: P92	7.83	4.84	5.09e+05	A/G2.H8	A/G2.H3'	1.58	5.00
89: P93	7.33	4.83	6.2e+05	A/T3.H6	A/G2.H3'	1.58	5.00
90: P94	7.33	4.80	1.1e+06	A/T3.H6	A/T3.H3'	1.58	3.00
91: P95	7.19	4.79	6.03e+05	B/C9.H6	B/G8.H3'	1.58	5.00
92: P96	7.19	4.34	1.08e+06	B/C9.H6	B/C9.H3'	1.58	3.00
93: P97	7.16	4.34	1.26e+06	B/T10.H6	B/T10.H3'	1.58	3.00
94: P98	7.06	4.34	7.65e+05	B/T11.H6	B/T10.H3'	1.58	5.00
95: P99	7.06	4.39	2.21e+06	B/T11.H6	B/T11.H3'	1.58	3.00
96: P100	8.37	4.40	1.7e+05	B/A12.H8	B/T11.H3'	1.58	5.00
97: P101	8.37	4.83	1.39e+06	B/A12.H8	B/A12.H3'	1.58	3.00
98: P102	7.20	4.83	4.35e+05	B/C13.H6	B/A12.H3'	1.58	5.00
99: P103	7.20	4.49	2.92e+06	B/C13.H6	B/C13.H3'	1.58	3.00
100: P104	7.80	4.49	4.19e+05	B/G14.H8	B/C13.H3'	1.58	5.00
101: P105	7.80	4.59	2.18e+06	B/G14.H8	B/G14.H3'	1.58	3.00
102: P106	8.22	4.80	6.51e+05	A/A4.H8	A/T3.H3'	1.58	5.00
103: P107	8.22	4.91	1.66e+06	A/A4.H8	A/A4.H3'	1.58	3.00
104: P108	7.65	4.91	9.83e+05	A/A5.H8	A/A4.H3'	1.58	5.00
105: P109	7.65	4.45	1.52e+06	A/A5.H8	A/A5.H3'	1.58	3.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
106: P110	6.82	4.45	1.76e+05	A/G6.H8	A/A5.H3'	1.58	5.00
107: P111	6.82	4.46	1.39e+05	A/G6.H8	A/G6.H3'	1.58	5.00
108: P112	6.74	4.47	4.16e+05	A/C7.H6	A/G6.H3'	1.58	5.00
109: P113	6.74	4.50	6.66e+05	A/C7.H6	A/C7.H3'	1.58	5.00
110: P133	7.79	4.03	5.24e+05	B/G14.H8	B/G14.H5'2	1.58	5.00
111: P135	7.64	3.67	2.18e+06	A/A5.H8	A/A5.H5'2	1.58	3.00
112: P136	7.65	3.03	4.09e+05	A/A5.H8	A/A5.H5'1	1.58	5.00
113: P137	5.59	1.80	3.92e+06	A/C1.H1'	A/C1.H2'2	1.58	3.00
114: P138	5.59	2.21	7.49e+06	A/C1.H1'	A/C1.H2'1	1.58	3.00
115: P139	5.13	2.20	4.87e+06	A/T3.H1'	A/T3.H2'2	1.58	3.00
116: P140	5.13	2.37	7.13e+06	A/T3.H1'	A/T3.H2'1	1.58	3.00
117: P141	6.09	2.57	4.61e+06	A/A4.H1'	A/A4.H2'2	1.58	3.00
118: P142	6.09	2.76	2.56e+06	A/A4.H1'	A/A4.H2'1	1.58	3.00
119: P143	5.69	2.09	5.01e+06	A/A5.H1'	A/A5.H2'2	1.58	3.00
120: P145	5.12	1.20	3.31e+06	A/C7.H1'	A/C7.H2'2	1.58	3.00
121: P146	5.12	1.71	6.42e+06	A/C7.H1'	A/C7.H2'1	1.58	3.00
122: P147	5.85	2.41	5.59e+06	B/G8.H1'	B/G8.H2'2	1.58	3.00
123: P148	5.84	2.65	5.67e+06	B/G8.H1'	B/G8.H2'1	1.58	3.00
124: P149	5.82	1.83	4.84e+06	B/C9.H1'	B/C9.H2'2	1.58	3.00
125: P150	5.81	2.16	7.38e+06	B/C9.H1'	B/C9.H2'1	1.58	3.00
126: P151	5.75	1.73	4.57e+06	B/T10.H1'	B/T10.H2'2	1.58	3.00
127: P152	5.75	2.27	4.43e+06	B/T10.H1'	B/T10.H2'1	1.58	3.00
128: P153	5.09	1.52	7.05e+06	B/T11.H1'	B/T11.H2'2	1.58	3.00
129: P154	5.10	1.62	1.96e+06	B/T11.H1'	B/T11.H2'1	1.58	3.00
130: P155	5.75	2.54	7.06e+06	B/A12.H1'	B/A12.H2'2	1.58	3.00
131: P157	5.72	2.50	2.09e+06	A/G2.H1'	A/G2.H2'1	1.58	3.00
132: P158	5.72	2.45	2.14e+06	A/G2.H1'	A/G2.H2'2	1.58	3.00
133: P159	6.00	2.22	7.55e+06	B/G14.H1'	B/G14.H2'2	1.58	3.00
134: P160	6.00	2.53	5.97e+06	B/G14.H1'	B/G14.H2'1	1.58	3.00
135: P161	5.24	2.01	6.16e+06	B/C13.H1'	B/C13.H2'1	1.58	3.00
136: P162	5.24	1.58	2.96e+06	B/C13.H1'	B/C13.H2'2	1.58	3.00
137: P163	5.75	2.81	1.17e+06	B/A12.H1'	B/A12.H2'1	1.58	3.00
138: P164	5.69	2.48	9.54e+06	A/A5.H1'	A/A5.H2'1	1.58	3.00
139: P165	4.79	3.45	2.16e+06	B/G8.H3'	B/G8.H5'1	1.58	3.00
140: P166	4.79	3.79	3.99e+06	B/G8.H3'	B/G8.H5'2	1.58	3.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
141: P167	4.83	3.79	5.54e+06	B/A12.H3'	B/A12.H5'1	1.58	3.00
142: P168	4.83	4.18	1.02e+07	B/A12.H3'	B/A12.H5'2	1.58	3.00
143: P169	4.59	3.76	6.09e+06	B/G14.H3'	B/G14.H5'1	1.58	3.00
144: P170	4.59	4.03	9.43e+06	B/G14.H3'	B/G14.H5'2	1.58	3.00
145: P171	4.57	3.55	7.16e+06	A/C1.H3'	A/C1.H5'1	1.58	3.00
146: P172	4.57	3.92	3.37e+06	A/C1.H3'	A/C1.H5'2	1.58	3.00
147: P173	4.50	3.80	5.81e+06	A/C7.H3'	A/C7.H5'2	1.58	3.00
148: P174	4.50	3.46	5.01e+06	A/C7.H3'	A/C7.H5'1	1.58	3.00
149: P175	4.49	3.35	5.57e+06	B/C13.H3'	B/C13.H5'1	1.58	3.00
150: P176	4.49	3.67	5.18e+06	B/C13.H3'	B/C13.H5'2	1.58	3.00
151: P177	4.45	3.02	4.81e+06	A/A5.H3'	A/A5.H5'1	1.58	3.00
152: P178	4.34	3.30	6.19e+06	B/T10.H3'	B/T10.H5'2	1.58	3.00
153: P179	4.34	2.79	2.47e+06	B/T10.H3'	B/T10.H5'1	1.58	3.00
154: P180	4.80	3.98	6.73e+06	A/T3.H3'	A/T3.H5'1	1.58	3.00
155: P181	4.80	4.02	2.31e+06	A/T3.H3'	A/T3.H5'2	1.58	3.00
156: P182	4.40	3.55	3.56e+06	B/T11.H3'	B/T11.H5'1	1.58	3.00
157: P183	4.39	3.71	5.75e+06	B/T11.H3'	B/T11.H5'2	1.58	3.00
158: P185	4.33	3.80	4.49e+06	B/C9.H3'	B/C9.H5'1	1.58	3.00
159: P186	4.92	4.11	4.54e+06	A/A4.H3'	A/A4.H5'2	1.58	3.00
160: P187	4.92	4.06	5.22e+06	A/A4.H3'	A/A4.H5'1	1.58	3.00
161: P188	4.85	3.95	2.45e+06	A/G2.H3'	A/G2.H5'2	1.58	3.00
162: P189	4.85	3.79	4.16e+06	A/G2.H3'	A/G2.H5'1	1.58	3.00
163: P190	4.45	3.66	2.6e+06	A/A5.H3'	A/A5.H5'2	1.58	3.00
164: P191	5.58	4.57	1.98e+06	A/C1.H1'	A/C1.H3'	1.58	3.00
165: P192	5.85	4.78	1.21e+06	B/G8.H1'	B/G8.H3'	1.58	3.00
166: P193	6.00	4.59	1.8e+06	B/G14.H1'	B/G14.H3'	1.58	3.00
167: P194	5.13	4.80	1.91e+06	A/T3.H1'	A/T3.H3'	1.58	3.00
168: P195	6.06	4.91	1.05e+06	A/A4.H1'	A/A4.H3'	1.58	3.00
169: P196	5.81	4.34	2.37e+06	B/C9.H1'	B/C9.H3'	1.58	3.00
170: P197	5.10	4.40	1.76e+06	B/T11.H1'	B/T11.H3'	1.58	3.00
171: P198	5.11	4.50	1.25e+06	A/C7.H1'	A/C7.H3'	1.58	3.00
172: P199	5.75	4.34	3.06e+06	B/T10.H1'	B/T10.H3'	1.58	3.00
173: P200	5.69	4.45	1.09e+06	A/A5.H1'	A/A5.H3'	1.58	3.00
174: P201	5.24	4.49	3.82e+06	B/C13.H1'	B/C13.H3'	1.58	3.00
175: P202	5.72	4.83	1.52e+06	A/G2.H1'	A/G2.H3'	1.58	3.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
176: P203	5.75	4.83	2.68e+06	B/A12.H1'	B/A12.H3'	1.58	3.00
177: P204	5.78	4.46	1.62e+06	A/G6.H1'	A/G6.H3'	1.58	3.00
178: P205	5.59	3.94	3.4e+06	A/C1.H1'	A/C1.H4'	1.58	3.00
179: P206	6.00	4.15	2.65e+06	B/G14.H1'	B/G14.H4'	1.58	3.00
180: P207	6.08	4.29	6.1e+05	A/A4.H1'	A/A4.H4'	1.58	5.00
181: P208	5.14	4.15	1.93e+06	A/T3.H1'	A/T3.H4'	1.58	3.00
182: P209	5.82	4.14	1.85e+06	B/C9.H1'	B/C9.H4'	1.58	3.00
183: P210	5.75	4.27	1.12e+06	B/A12.H1'	B/A12.H4'	1.58	3.00
184: P211	5.75	4.15	8.88e+05	B/T10.H1'	B/T10.H4'	1.58	5.00
185: P212	5.11	4.10	1.12e+06	A/C7.H1'	A/C7.H4'	1.58	3.00
186: P213	5.72	3.96	8.61e+05	A/G2.H1'	A/G2.H4'	1.58	5.00
187: P214	5.85	4.08	1.64e+06	B/G8.H1'	B/G8.H4'	1.58	3.00
188: P215	4.57	1.80	7.16e+06	A/C1.H3'	A/C1.H2'2	1.58	3.00
189: P216	4.57	2.20	3.86e+06	A/C1.H3'	A/C1.H2'1	1.58	3.00
190: P217	4.80	2.19	4.81e+06	A/T3.H3'	A/T3.H2'2	1.58	3.00
191: P218	4.80	2.37	7.56e+06	A/T3.H3'	A/T3.H2'1	1.58	3.00
192: P219	4.92	2.57	6.47e+06	A/A4.H3'	A/A4.H2'2	1.58	3.00
193: P220	4.92	2.77	3.11e+06	A/A4.H3'	A/A4.H2'1	1.58	3.00
194: P221	4.51	1.20	2.51e+06	A/C7.H3'	A/C7.H2'2	1.58	3.00
195: P222	4.51	1.72	2.42e+06	A/C7.H3'	A/C7.H2'1	1.58	3.00
196: P225	4.60	2.22	5.68e+06	B/G14.H3'	B/G14.H2'2	1.58	3.00
197: P226	4.60	2.53	4.58e+06	B/G14.H3'	B/G14.H2'1	1.58	3.00
198: P227	4.83	2.82	2.91e+06	B/A12.H3'	B/A12.H2'1	1.58	3.00
199: P228	4.83	2.54	6.05e+06	B/A12.H3'	B/A12.H2'2	1.58	3.00
200: P229	4.85	2.47	8.28e+06	A/G2.H3'	A/G2.H2'2	1.58	3.00
201: P230	4.85	2.51	7.09e+06	A/G2.H3'	A/G2.H2'1	1.58	3.00
202: P231	4.79	2.65	1.27e+06	B/G8.H3'	B/G8.H2'1	1.58	3.00
203: P232	4.79	2.41	2e+06	B/G8.H3'	B/G8.H2'2	1.58	3.00
204: P233	4.49	1.58	3.96e+06	B/C13.H3'	B/C13.H2'2	1.58	3.00
205: P234	4.49	2.02	1.21e+07	B/C13.H3'	B/C13.H2'1	1.58	3.00
206: P235	4.34	1.82	5.2e+06	B/C9.H3'	B/C9.H2'2	1.58	3.00
207: P236	4.34	2.17	4.25e+06	B/C9.H3'	B/C9.H2'1	1.58	3.00
208: P238	4.39	1.53	5.71e+06	B/T11.H3'	B/T11.H2'2	1.58	3.00
209: P239	4.36	1.72	4.29e+06	B/T10.H3'	B/T10.H2'2	1.58	3.00
210: P240	4.36	2.27	1.58e+06	B/T10.H3'	B/T10.H2'1	1.58	3.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
211: P241	4.45	2.09	4.43e+06	A/A5.H3'	A/A5.H2'2	1.58	3.00
212: P242	4.45	2.48	5.54e+06	A/A5.H3'	A/A5.H2'1	1.58	3.00
213: P244	2.02	1.58	4.21e+06	B/C13.H2'1	B/C13.H2'2	1.58	3.00
214: P245	1.72	1.20	2.41e+06	A/C7.H2'1	A/C7.H2'2	1.58	3.00
215: P246	2.53	2.23	2.21e+07	B/G14.H2'1	B/G14.H2'2	1.58	3.00
216: P247	2.77	2.57	6.53e+06	A/A4.H2'1	A/A4.H2'2	1.58	3.00
217: P248	2.83	2.54	2.25e+06	B/A12.H2'1	B/A12.H2'2	1.58	3.00
218: P249	2.20	1.80	1.85e+07	A/C1.H2'1	A/C1.H2'2	1.58	3.00
219: P251	2.18	1.82	3.83e+06	B/C9.H2'1	B/C9.H2'2	1.58	3.00
220: P252	2.37	2.20	2.88e+06	A/T3.H2'1	A/T3.H2'2	1.58	3.00
221: P253	2.67	2.42	2.16e+06	B/G8.H2'1	B/G8.H2'2	1.58	3.00
222: P255	3.02	2.48	4.6e+06	A/A5.H5'1	A/A5.H2'1	1.58	3.00
223: P256	3.35	2.01	3.31e+06	B/C13.H5'1	B/C13.H2'1	1.58	3.00
224: P257	3.30	2.78	2.05e+06	B/T10.H5'2	B/T10.H5'1	1.58	3.00
225: P258	4.31	2.09	1.61e+06	A/A5.H4'	A/A5.H2'2	1.58	3.00
226: P259	4.31	2.48	4.84e+06	A/A5.H4'	A/A5.H2'1	1.58	3.00
227: P260	4.29	2.77	6.43e+05	A/A4.H4'	A/A4.H2'1	1.58	5.00
228: P261	4.29	2.56	4.08e+06	A/A4.H4'	A/A4.H2'2	1.58	3.00
229: P262	4.27	2.81	7.63e+05	B/A12.H4'	B/A12.H2'1	1.58	5.00
230: P265	3.93	2.20	5.47e+06	A/C1.H4'	A/C1.H2'1	1.58	3.00
231: P267	6.73	4.60	2.03e+06	A/C7.H6	A/C7.H5	1.58	3.00
232: P268	7.21	5.54	5.04e+06	B/C13.H6	B/C13.H5	1.58	3.00
233: P269	7.51	5.74	1.24e+07	A/C1.H6	A/C1.H5	1.58	3.00
234: P270	7.20	5.22	6.55e+06	B/C9.H6	B/C9.H5	1.58	3.00
235: P271	7.97	7.81	5.7e+05	B/A12.H2	A/A4.H2	1.58	5.00
236: P276	7.95	6.06	6.09e+05	B/A12.H2	A/A4.H1'	1.58	5.00
237: P277	7.81	6.06	3.5e+05	A/A4.H2	A/A4.H1'	1.58	5.00
238: P278	7.81	5.10	5.14e+05	A/A4.H2	B/T11.H1'	1.58	5.00
239: P272	7.95	5.75	1.26e+06	B/A12.H2	B/A12.H1'	1.58	3.00
240: P274	7.95	5.13	1.95e+06	B/A12.H2	A/T3.H1'	1.58	3.00
241: P275	7.86	5.75	2.42e+05	A/A5.H2	B/T10.H1'	1.58	5.00
242: P279	7.81	5.75	1.25e+05	A/A4.H2	B/A12.H1'	1.58	5.00
243: P280	7.06	1.45	2.44e+06	B/T11.H6	B/T11.C5M	1.58	3.00
244: P281	7.16	1.45	7.87e+05	B/T10.H6	B/T11.C5M	1.58	5.00
245: P283	7.16	1.37	2.93e+06	B/T10.H6	B/T10.C5M	1.58	3.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
246: P284	7.83	1.56	1.47e+06	A/G2.H8	A/T3.C5M	1.58	3.00
247: P285	7.33	1.56	2.34e+06	A/T3.H6	A/T3.C5M	1.58	3.00
248: P286	5.82	1.36	6.97e+05	B/C9.H1'	B/T10.C5M	1.58	5.00
249: P287	5.22	1.36	7.84e+05	B/C9.H5	B/T10.C5M	1.58	5.00
250: P288	5.75	1.45	7.97e+05	B/T10.H1'	B/T11.C5M	1.58	5.00
251: P289	5.10	1.45	3.05e+05	B/T11.H1'	B/T11.C5M	1.58	5.00
252: P290	5.72	1.56	5.84e+05	A/G2.H1'	A/T3.C5M	1.58	5.00
253: P291	5.14	1.56	4.7e+05	A/T3.H1'	A/T3.C5M	1.58	5.00
254: P293	4.36	1.45	6.96e+05	B/T10.H3'	B/T11.C5M	1.58	5.00
255: P294	4.85	1.56	5.4e+05	A/G2.H3'	A/T3.C5M	1.58	5.00
256: P295	2.18	1.37	1.03e+06	B/C9.H2'1	B/T10.C5M	1.58	3.00
257: P296	2.27	1.45	1.28e+06	B/T10.H2'1	B/T11.C5M	1.58	3.00
258: P297	1.84	1.37	1.25e+06	B/C9.H2'2	B/T10.C5M	1.58	3.00
259: P298	1.73	1.45	1.52e+06	B/T10.H2'2	B/T11.C5M	1.58	3.00
260: P299	2.51	1.56	2.18e+06	A/G2.H2'1	A/T3.C5M	1.58	3.00
261: P300	2.46	1.56	1.78e+06	A/G2.H2'2	A/T3.C5M	1.58	3.00
262: P301	7.74	7.37	5.56e+06	A/OH1.H7'2	A/OH1.H6'2	1.58	3.00
263: P302	7.50	7.37	5.5e+06	A/OH1.H5'2	A/OH1.H6'2	1.58	3.00
264: P303	8.24	7.63	6.46e+06	A/OH1.H6'1	A/OH1.H7'1	1.58	3.00
265: P304	7.94	7.49	4.96e+06	A/OH1.H4'2	A/OH1.H5'2	1.58	3.00
266: P306	7.95	6.96	2.94e+06	B/A12.H2	A/OH1.H6	1.58	3.00
267: P307	7.28	7.63	7.13e+05	A/OH1.H3'1	A/OH1.H7'1	1.58	5.00
268: P308	7.81	7.26	8.14e+05	A/A4.H2	A/OH1.H3'1	1.58	5.00
269: P310	7.93	7.37	6.62e+05	A/OH1.H4'2	A/OH1.H6'2	1.58	5.00
270: P313	7.81	7.28	4.09e+05	A/OH1.H4'1	A/OH1.H3'1	1.58	5.00
271: P314	7.28	7.86	1.43e+05	A/OH1.H3'1	A/A5.H2	1.58	5.00
272: P315	7.74	7.48	7.65e+05	A/OH1.H7'2	A/OH1.H5'2	1.58	5.00
273: P316	7.93	7.65	1.62e+05	A/OH1.H4'2	A/OH1.H3'2	1.58	5.00
274: P317	7.80	7.62	2.77e+05	A/OH1.H4'1	A/OH1.H7'1	1.58	5.00
275: P319	7.81	8.24	7.25e+05	A/OH1.H4'1	A/OH1.H6'1	1.58	5.00
276: P322	8.36	5.53	9.34e+05	B/A12.H8	B/C13.H5	1.58	5.00
277: P323	7.50	5.82	3.19e+06	A/OH1.H5'2	B/C9.H1'	1.58	3.00
278: P324	7.58	5.21	4.85e+05	B/G8.H8	B/C9.H5	1.58	5.00
279: P325	7.99	5.16	2.28e+06	A/OH1.H3	A/OH1.H8B	1.58	3.00
280: P326	7.99	5.75	1.14e+06	A/OH1.H3	B/A12.H1'	1.58	3.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
281: P327	7.93	5.82	2.51e+06	A/OH1.H4'2	B/C9.H1'	1.58	3.00
282: P328	7.93	5.83	1.08e+06	A/OH1.H4'2	B/G8.H1'	1.58	3.00
283: P338	7.39	5.82	4.36e+05	A/OH1.H6'2	B/C9.H1'	1.58	5.00
284: P339	6.82	4.60	1.36e+05	A/G6.H8	A/C7.H5	1.58	5.00
285: P341	7.59	4.79	7e+05	B/G8.H8	B/G8.H3'	1.58	5.00
286: P347	7.81	4.50	3.7e+05	A/A4.H2	A/OH1.H8A2	1.58	5.00
287: P350	8.24	4.10	2.95e+05	A/OH1.H6'1	A/C7.H4'	1.58	5.00
288: P353	7.50	4.03	2.31e+05	A/OH1.H5'2	B/C9.H5'2	1.58	5.00
289: P354	7.50	4.14	1.41e+06	A/OH1.H5'2	B/C9.H4'	1.58	3.00
290: P359	7.80	4.00	3.09e+05	A/OH1.H4'1	B/T11.H4'	1.58	5.00
291: P361	7.94	4.33	5.22e+05	A/OH1.H4'2	B/C9.H3'	1.58	5.00
292: P363	8.24	3.80	1.35e+06	A/OH1.H6'1	A/C7.H5'2	1.58	3.00
293: P364	8.24	3.46	1.34e+06	A/OH1.H6'1	A/C7.H5'1	1.58	3.00
294: P367	7.51	3.55	2.53e+06	A/C1.H6	A/C1.H5'1	1.58	3.00
295: P371	7.65	3.47	2.65e+05	A/OH1.H3'2	A/C7.H5'1	1.58	5.00
296: P372	7.63	3.79	7.13e+05	A/OH1.H7'1	B/A12.H5'1	1.58	5.00
297: P374	7.50	3.31	1.09e+06	A/OH1.H5'2	B/T10.H5'2	1.58	3.00
298: P376	7.93	3.32	7.1e+05	A/OH1.H4'2	B/T10.H5'2	1.58	5.00
299: P379	7.50	2.79	7.73e+05	A/OH1.H5'2	B/T10.H5'1	1.58	5.00
300: P380	7.93	2.79	1.53e+06	A/OH1.H4'2	B/T10.H5'1	1.58	3.00
301: P382	7.99	2.54	6.05e+05	A/OH1.H3	A/OH1.C34	1.58	5.00
302: P383	7.95	2.54	7.91e+05	B/A12.H2	A/OH1.C34	1.58	5.00
303: P384	6.98	2.55	5.32e+05	A/OH1.H6	A/OH1.C34	1.58	5.00
304: P386	7.65	1.72	4.91e+05	A/OH1.H3'2	B/T10.H2'2	1.58	5.00
305: P387	7.93	1.82	8.12e+05	A/OH1.H4'2	B/C9.H2'2	1.58	5.00
306: P389	7.93	2.18	2.87e+06	A/OH1.H4'2	B/C9.H2'1	1.58	3.00
307: P390	5.25	2.55	5.06e+05	B/C13.H1'	A/OH1.C34	1.58	5.00
308: P391	5.16	2.55	2.17e+06	A/OH1.H8B	A/OH1.C34	1.58	3.00
309: P392	5.23	2.41	8.39e+05	B/C9.H5	B/G8.H2'2	1.58	5.00
310: P393	5.23	2.66	1.05e+06	B/C9.H5	B/G8.H2'1	1.58	3.00
311: P395	5.55	1.57	1.01e+06	B/C13.H5	B/C13.H2'2	1.58	3.00
312: P396	5.22	1.83	1.26e+06	B/C9.H5	B/C9.H2'2	1.58	3.00
313: P397	5.56	2.82	1.88e+06	B/C13.H5	B/A12.H2'1	1.58	3.00
314: P398	5.56	2.54	2.97e+06	B/C13.H5	B/A12.H2'2	1.58	3.00
315: P399	4.62	1.19	3.63e+05	A/C7.H5	A/C7.H2'2	1.58	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
316: P402	6.98	5.06	9.77e+05	A/OH1.H6	A/OH1.H8A	1.58	5.00
317: P403	7.96	5.06	3.25e+06	B/A12.H2	A/OH1.H8A	1.58	3.00
318: P404	7.96	5.16	1.77e+06	B/A12.H2	A/OH1.H8B	1.58	3.00
319: P405	6.98	5.16	1.17e+06	A/OH1.H6	A/OH1.H8B	1.58	3.00
320: P406	5.16	4.50	3.01e+05	A/OH1.H8B	A/OH1.H8A2	1.58	5.00
321: P407	4.83	4.27	6.98e+06	B/A12.H3'	B/A12.H4'	1.58	3.00
322: P408	4.93	4.29	7.53e+06	A/A4.H3'	A/A4.H4'	1.58	3.00
323: P409	4.60	4.15	6.45e+06	B/G14.H3'	B/G14.H4'	1.58	3.00
324: P410	4.80	4.15	4.85e+06	A/T3.H3'	A/T3.H4'	1.58	3.00
325: P411	4.79	4.08	2.46e+06	B/G8.H3'	B/G8.H4'	1.58	3.00
326: P412	4.50	4.10	3.63e+05	A/C7.H3'	A/C7.H4'	1.58	5.00
327: P413	4.83	3.97	2.27e+06	A/G2.H3'	A/G2.H4'	1.58	3.00
328: P414	4.49	4.02	1.61e+06	B/C13.H3'	B/C13.H4'	1.58	3.00
329: P415	7.81	3.69	1.47e+05	A/A4.H2	A/OH1.H1A	1.58	5.00
330: P416	7.86	3.69	9.66e+05	A/A5.H2	A/OH1.H1A	1.58	5.00
331: P417	7.86	3.81	4.04e+05	A/A5.H2	A/OH1.H1B	1.58	5.00
332: P418	7.96	3.69	7.59e+05	B/A12.H2	A/OH1.H1A	1.58	5.00
333: P419	7.96	3.81	1.13e+06	B/A12.H2	A/OH1.H1B	1.58	3.00
334: P421	7.81	3.81	1.4e+06	A/A4.H2	A/OH1.H1B	1.58	3.00
335: P422	7.28	3.81	4.09e+05	A/OH1.H3'1	A/OH1.H1B	1.58	5.00
336: P423	7.27	3.69	4.24e+05	A/OH1.H3'1	A/OH1.H1A	1.58	5.00
337: P424	4.50	3.81	5.89e+05	A/OH1.H8A2	A/OH1.H1B	1.58	5.00
338: P425	4.50	3.69	6.48e+05	A/OH1.H8A2	A/OH1.H1A	1.58	5.00
339: P426	5.26	3.67	6.94e+05	B/C13.H1'	B/C13.H5'2	1.58	5.00
340: P427	5.59	3.55	1.08e+06	A/C1.H1'	A/C1.H5'1	1.58	3.00
341: P436	3.80	3.45	1.74e+07	A/C7.H5'2	A/C7.H5'1	1.58	3.00
342: P437	3.92	3.55	1.11e+07	A/C1.H5'2	A/C1.H5'1	1.58	3.00
343: P438	4.10	3.45	1.69e+06	A/C7.H4'	A/C7.H5'1	1.58	3.00
344: P439	4.39	3.99	1.25e+06	B/T11.H3'	B/T11.H4'	1.58	3.00

APPENDIX XIII

The 2D ^{1}H NOESY of the 5′-($C^1G^2A^3A^4A^5A^6A^7C^8G^9G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA duplex.

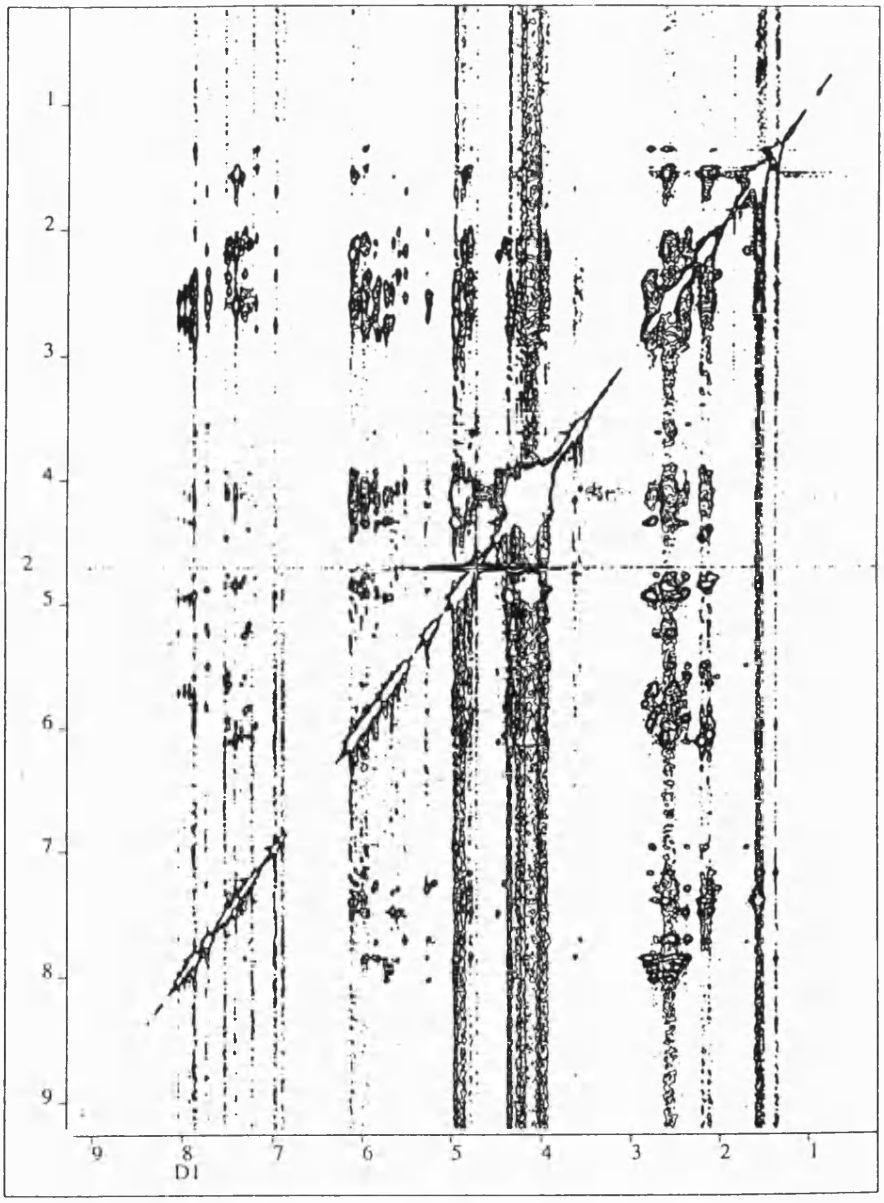


Figure 148 The 2D 600 MHz ^{1}H NOESY of the 5′-($C^1G^2A^3A^4A^5A^6A^7C^8G^9G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA duplex.

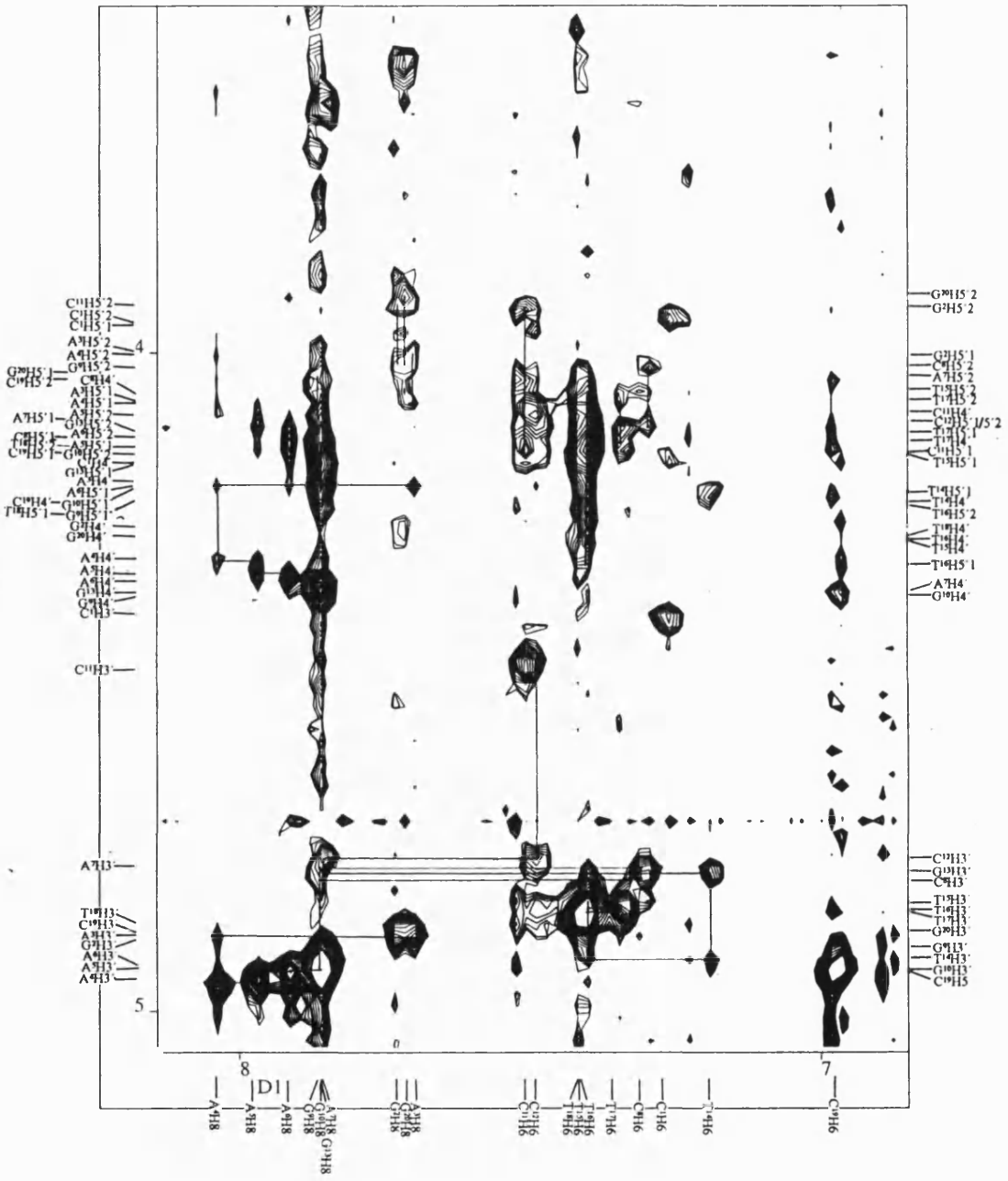


Figure 149 The 2D 600 MHz ^{1}H NOESY of the 5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA duplex. The aromatic→H3′/H4′/H5′1/H5′2 region.

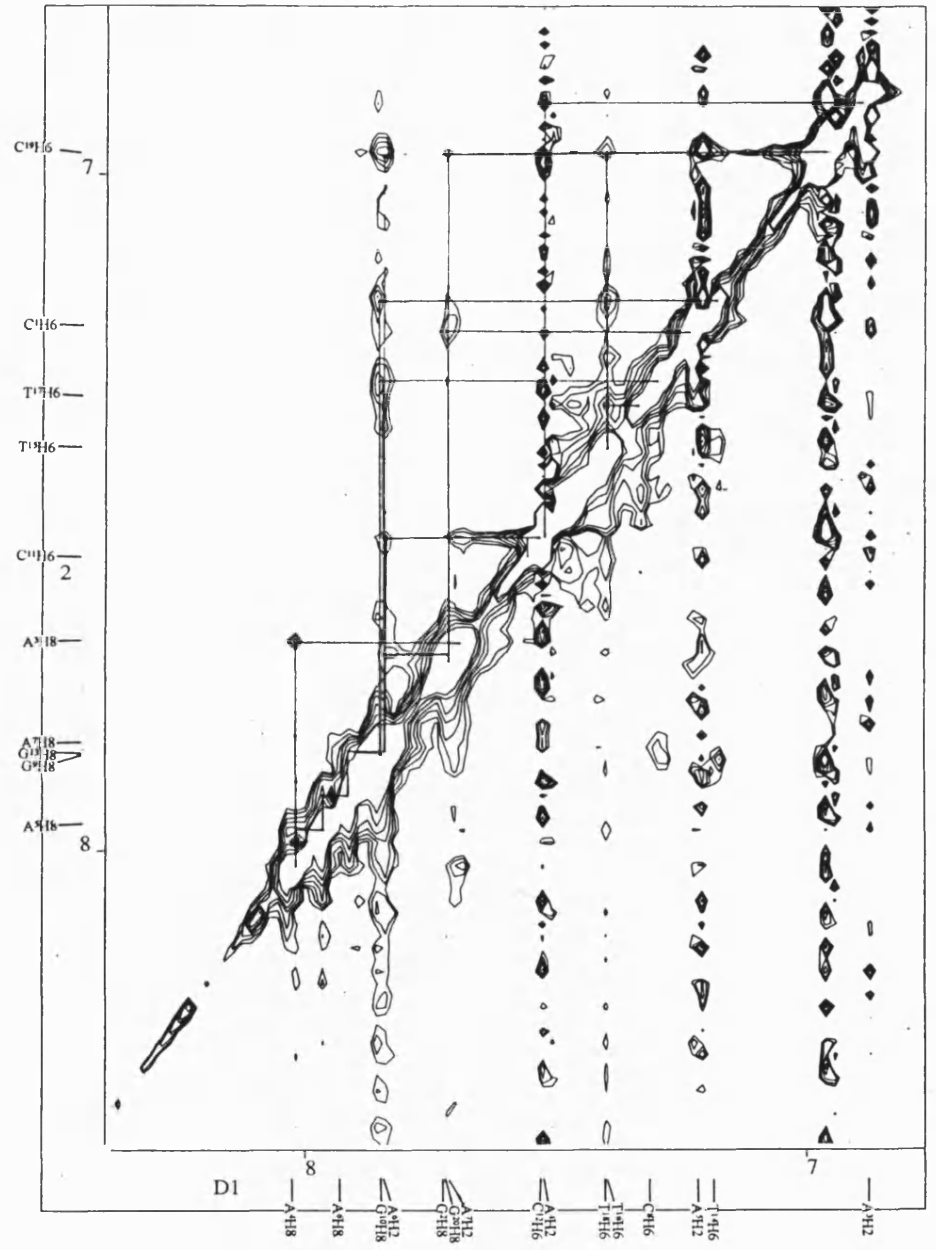


Figure 150 The 2D 600 MHz ^{1}H NOESY of the 5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA duplex. The aromatic→aromatic region.

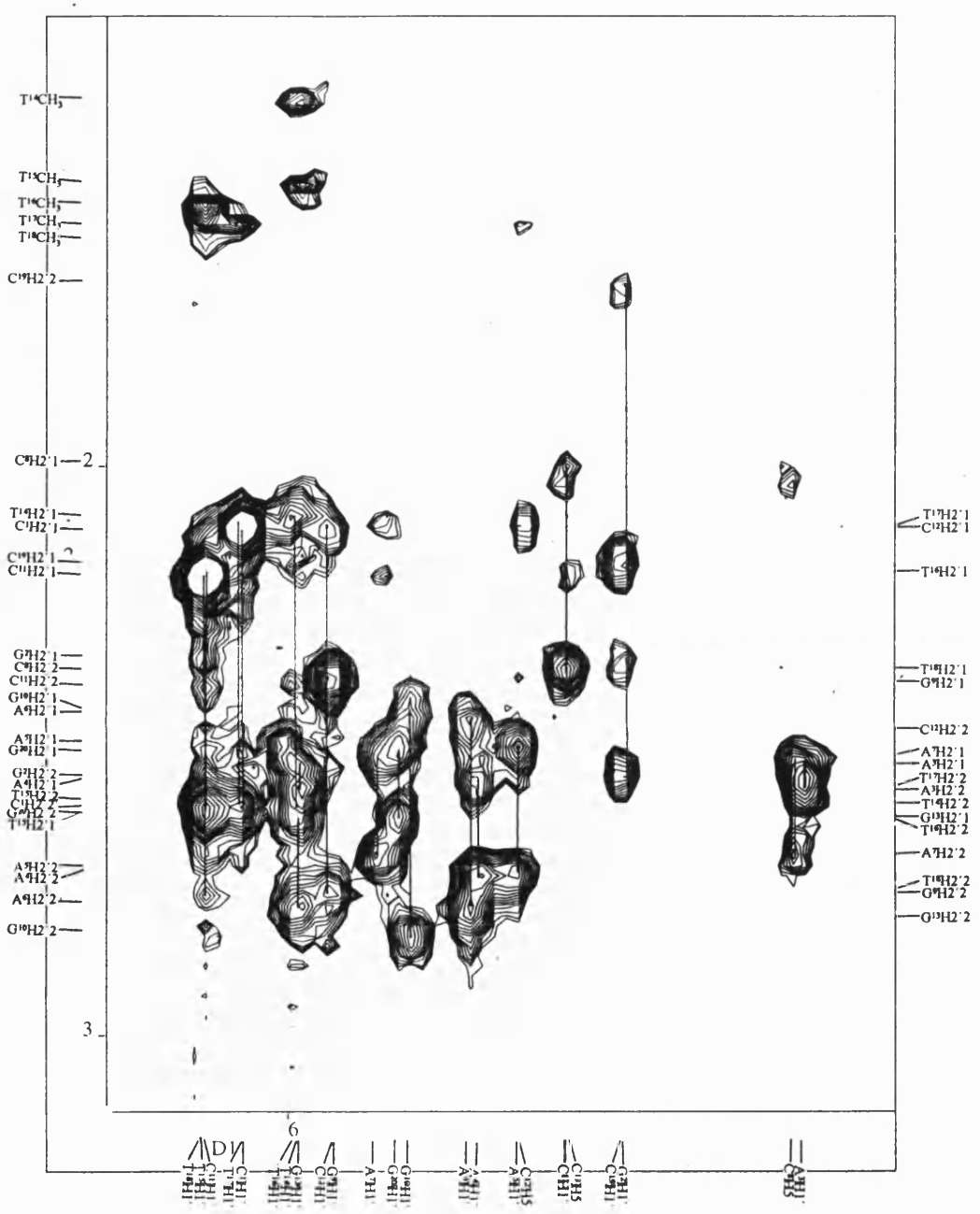


Figure 151 The 2D 600 MHz ^{1}H NOESY of the 5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA duplex. The H1′→H2′1/H2′2 region.

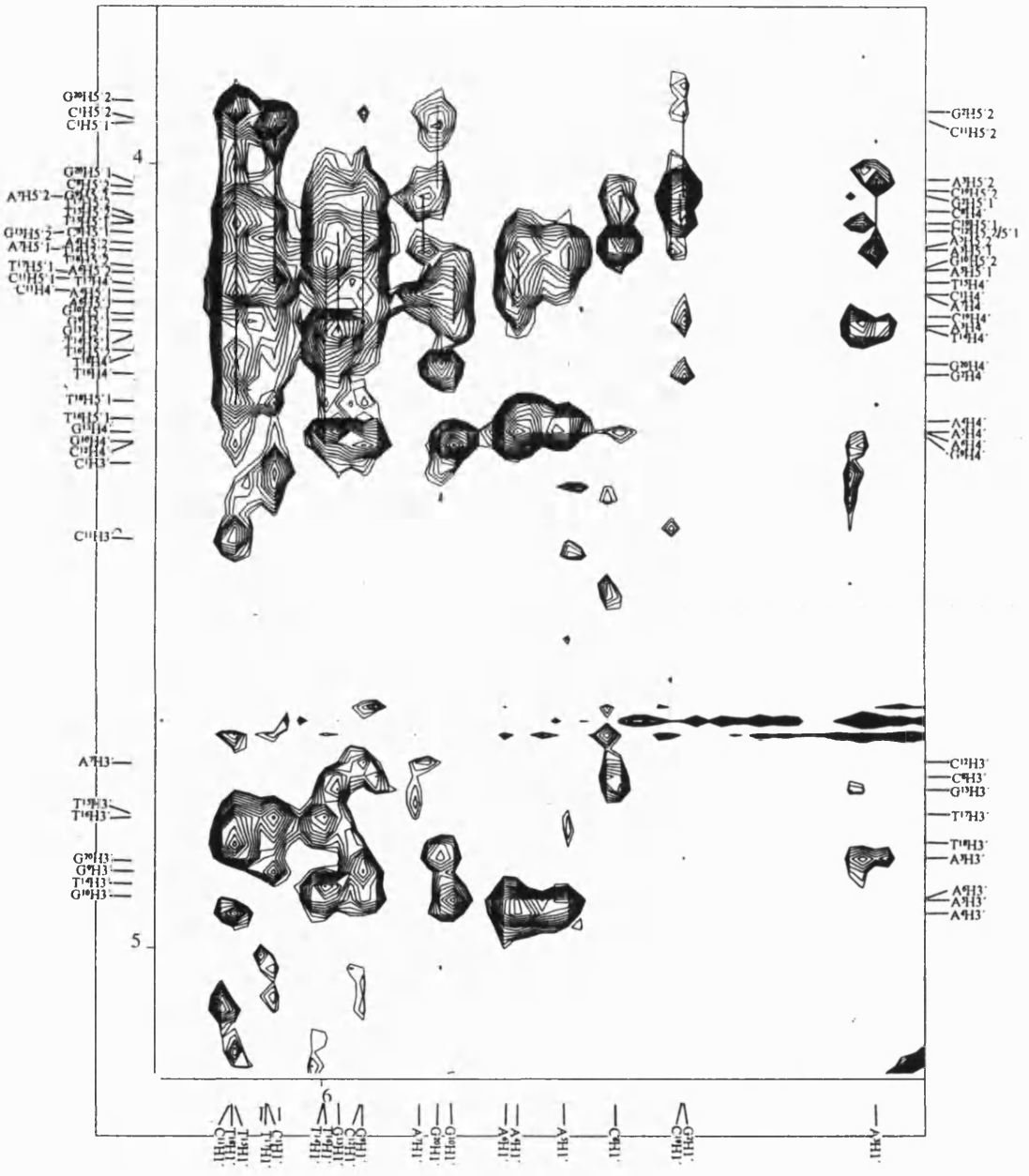


Figure 152 The 2D 600 MHz ^{1}H NOESY of the 5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA duplex. The H1′→H3′/H4′/H5′1/H5′2 NOE region.

APPENDIX XIV

The 2D 600 MHz ^{1}H DQF-COSY of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct.

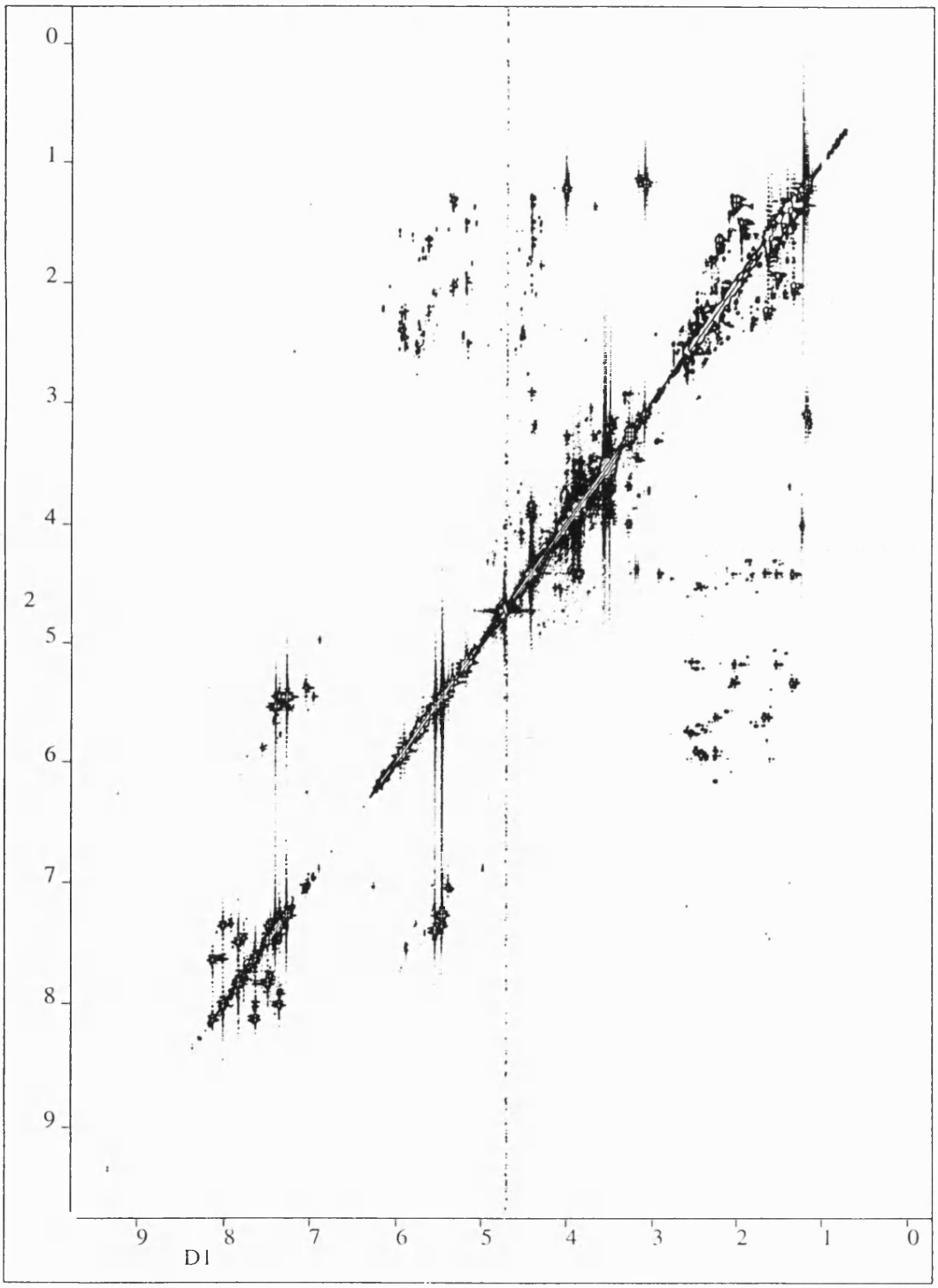


Figure 153 The 2D 600 MHz ^{1}H DQF-COSY of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct.

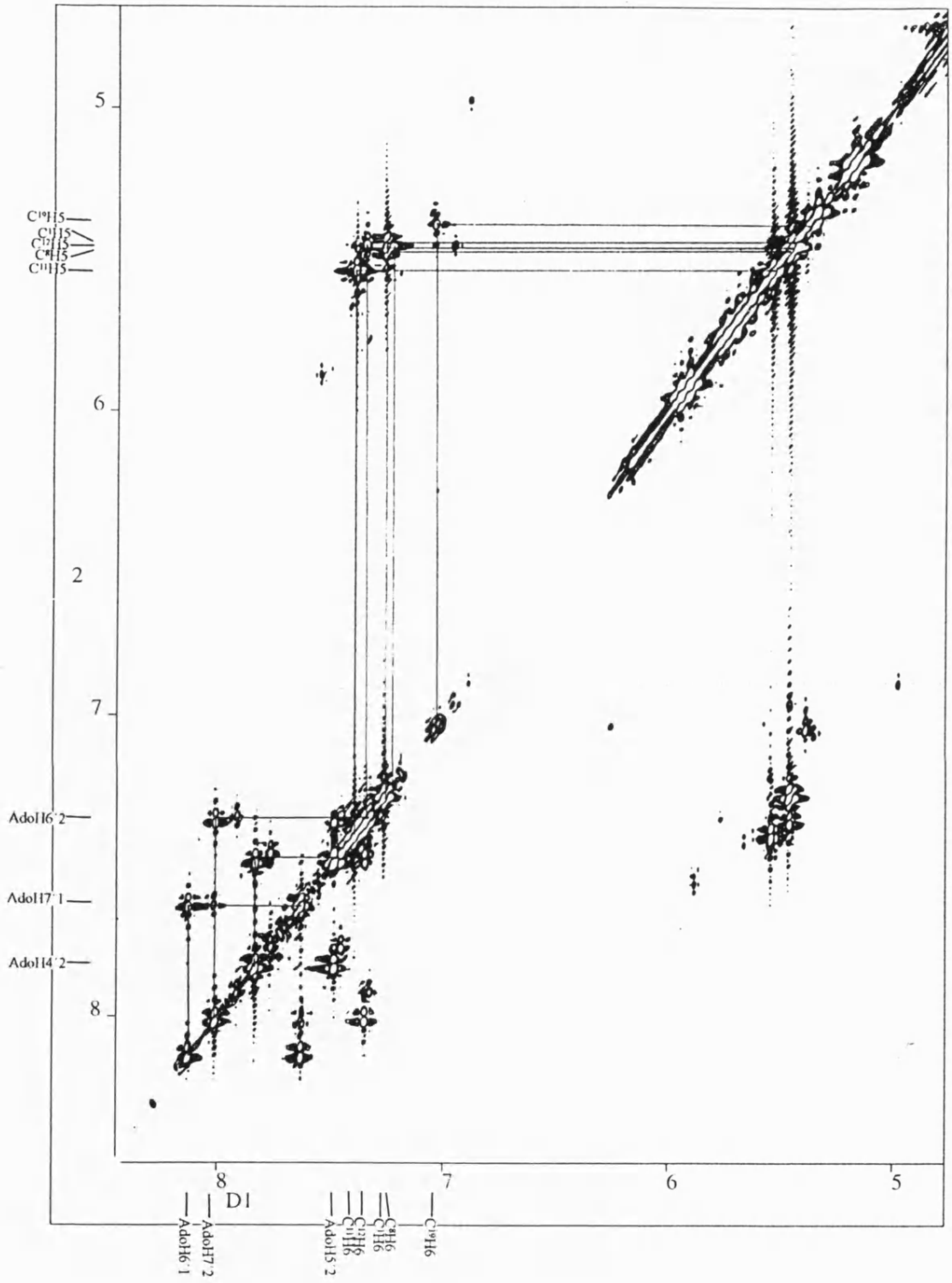


Figure 154 The 2D 600 MHz ^{1}H DQF-COSY of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The aromatics→aromatics/CH5 region.

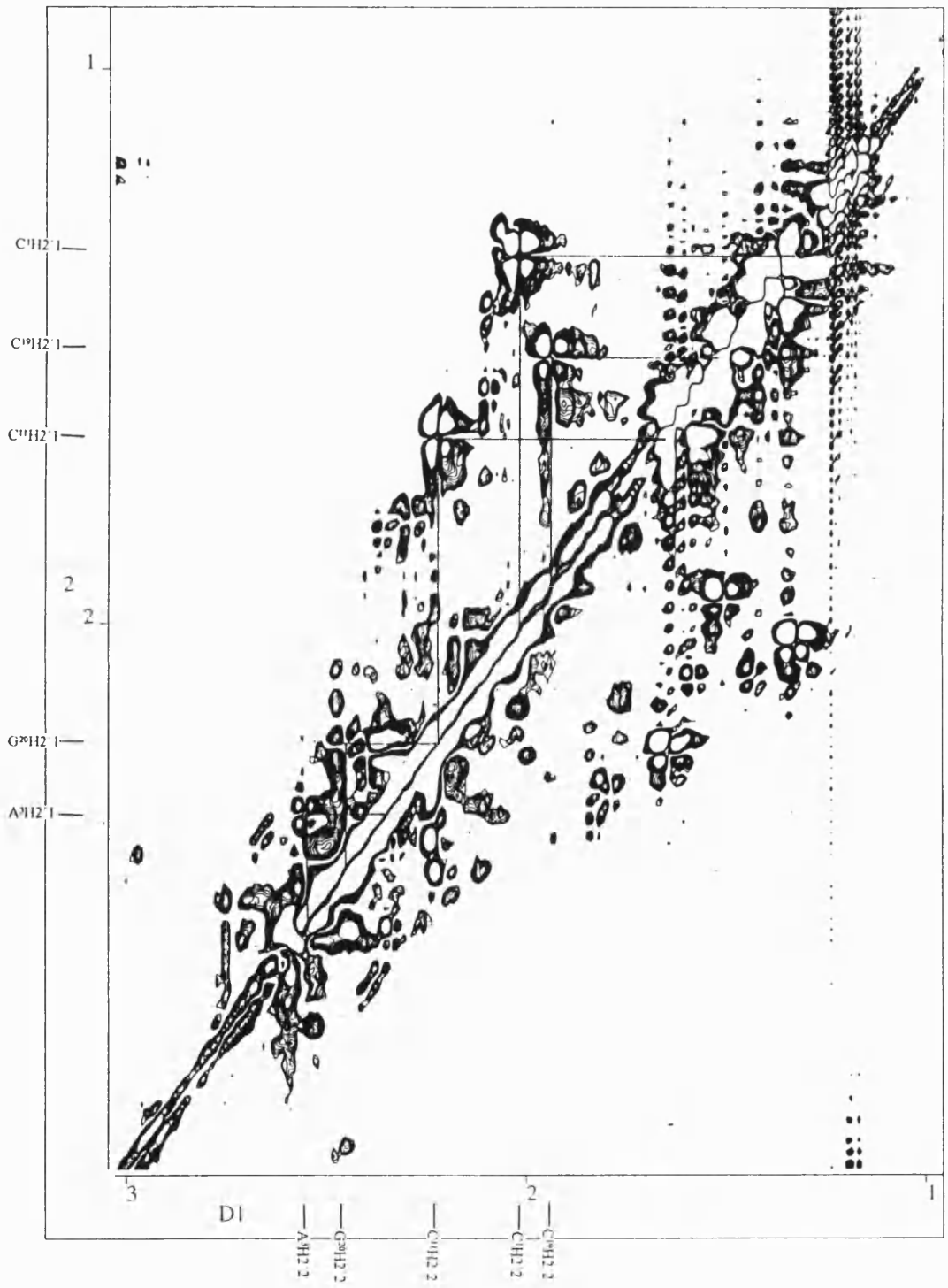


Figure 155 The 2D 600 MHz ^{1}H DQF-COSY of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The H2′1→H2′2 region.

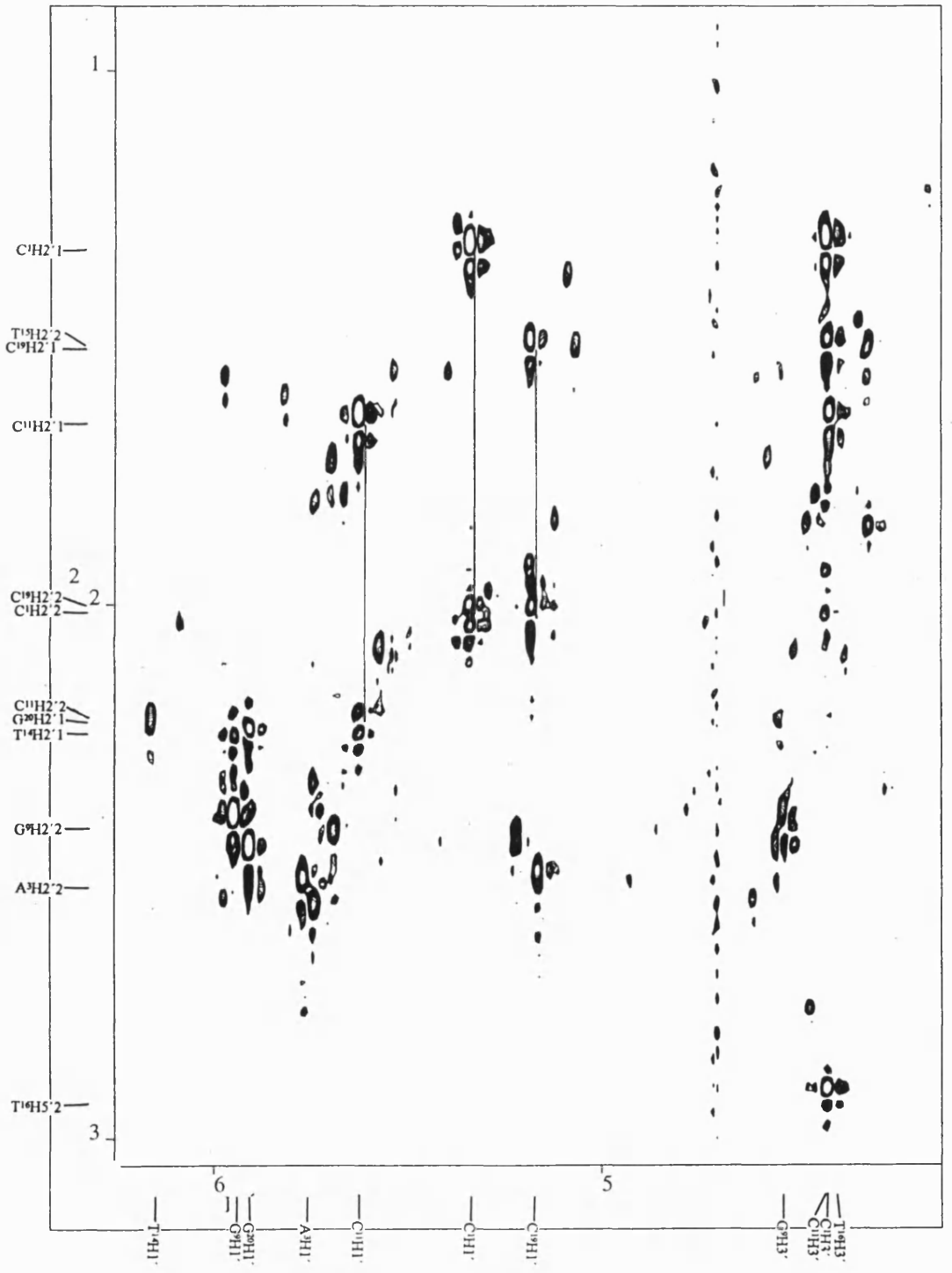


Figure 156 The 2D 600 MHz ^{1}H DQF-COSY of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The H1′→H2′1/H2′2 region.

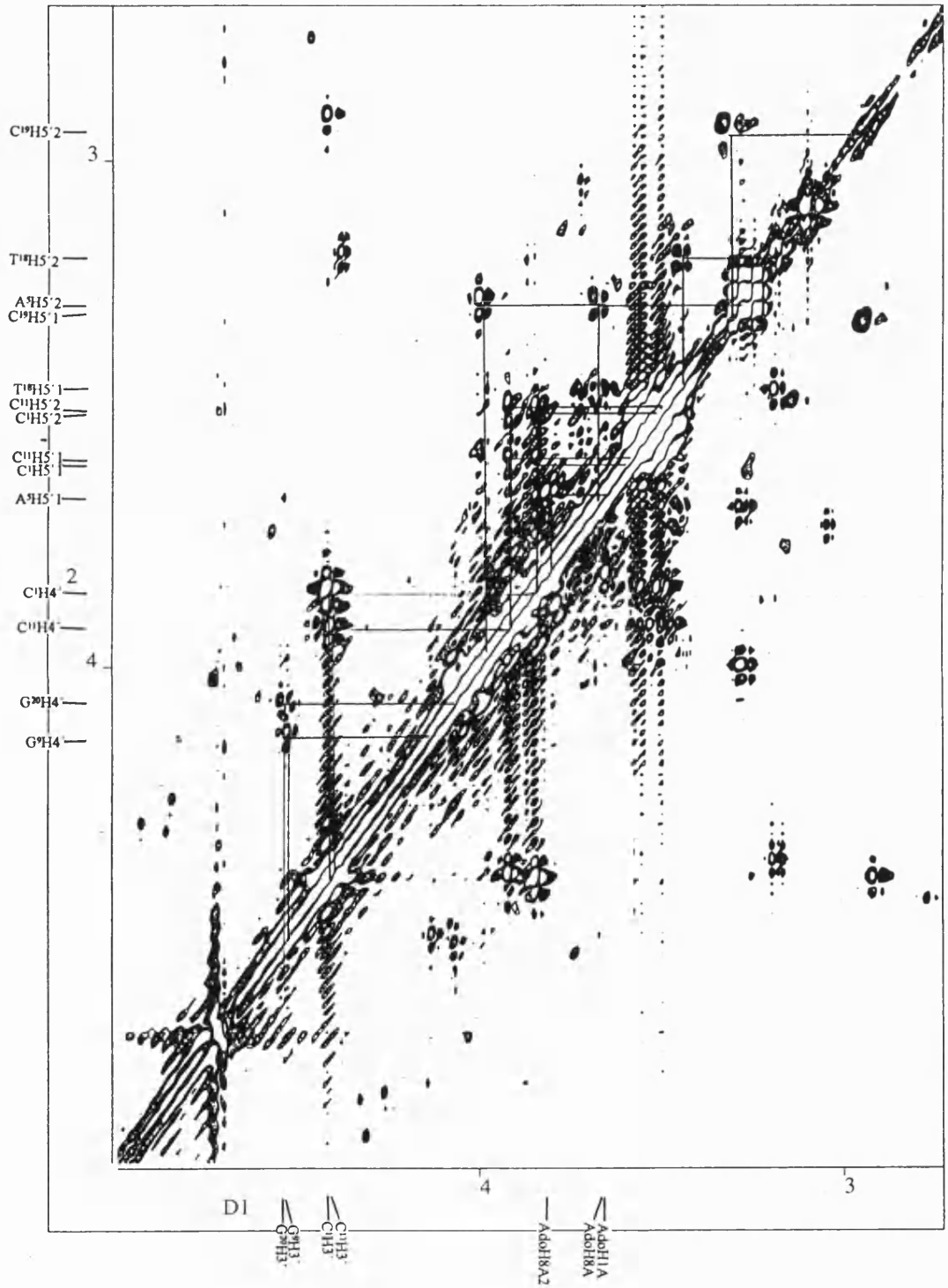


Figure 157 The 2D 600 MHz ^{1}H DQF-COSY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The H3′, H4′, H5′1 and H5′2 COSY connectivities.

APPENDIX XV

The 2D 600 MHz ^{1}H TOCSYspectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct.

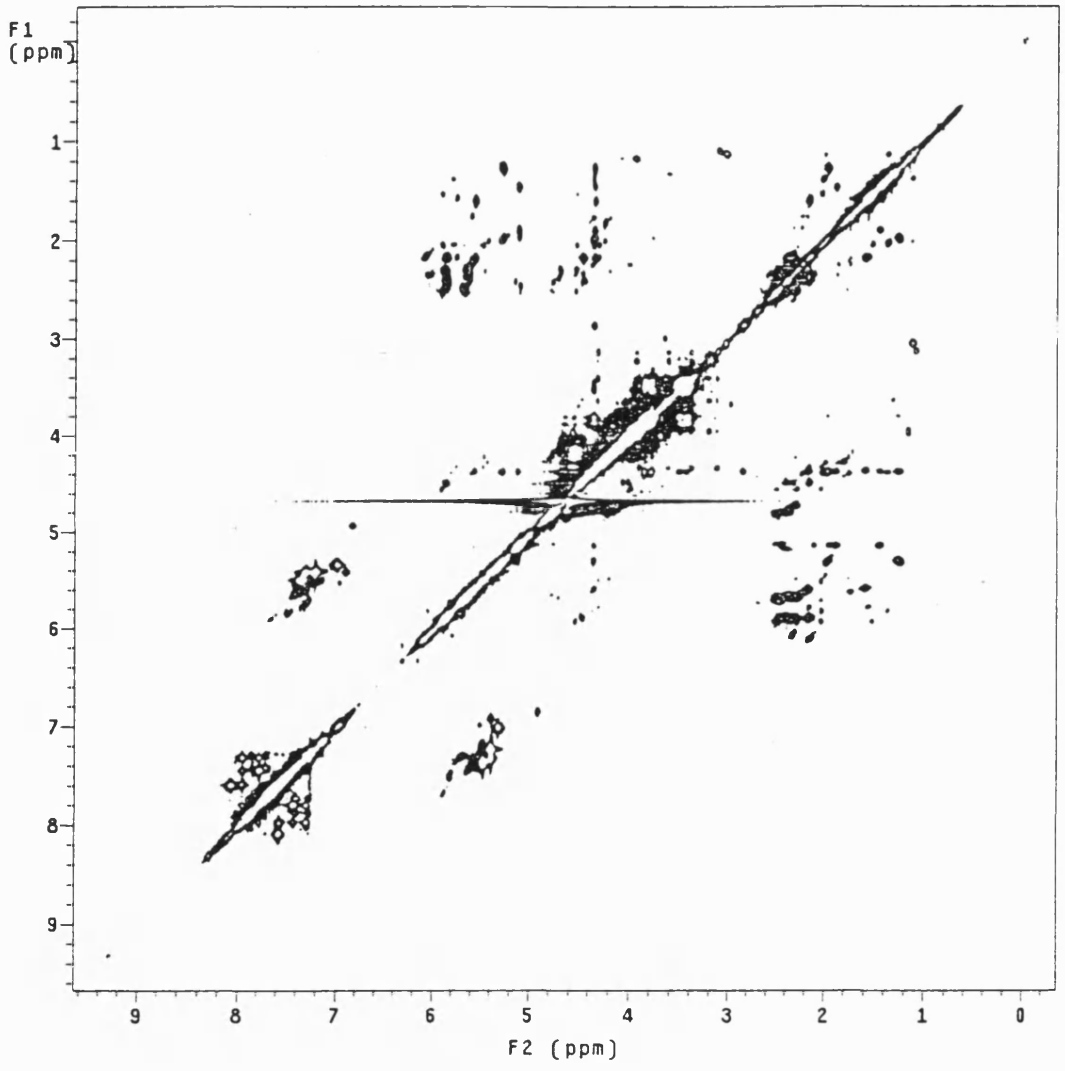


Figure 158 The 2D 600 MHz ^{1}H TOCSY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct.

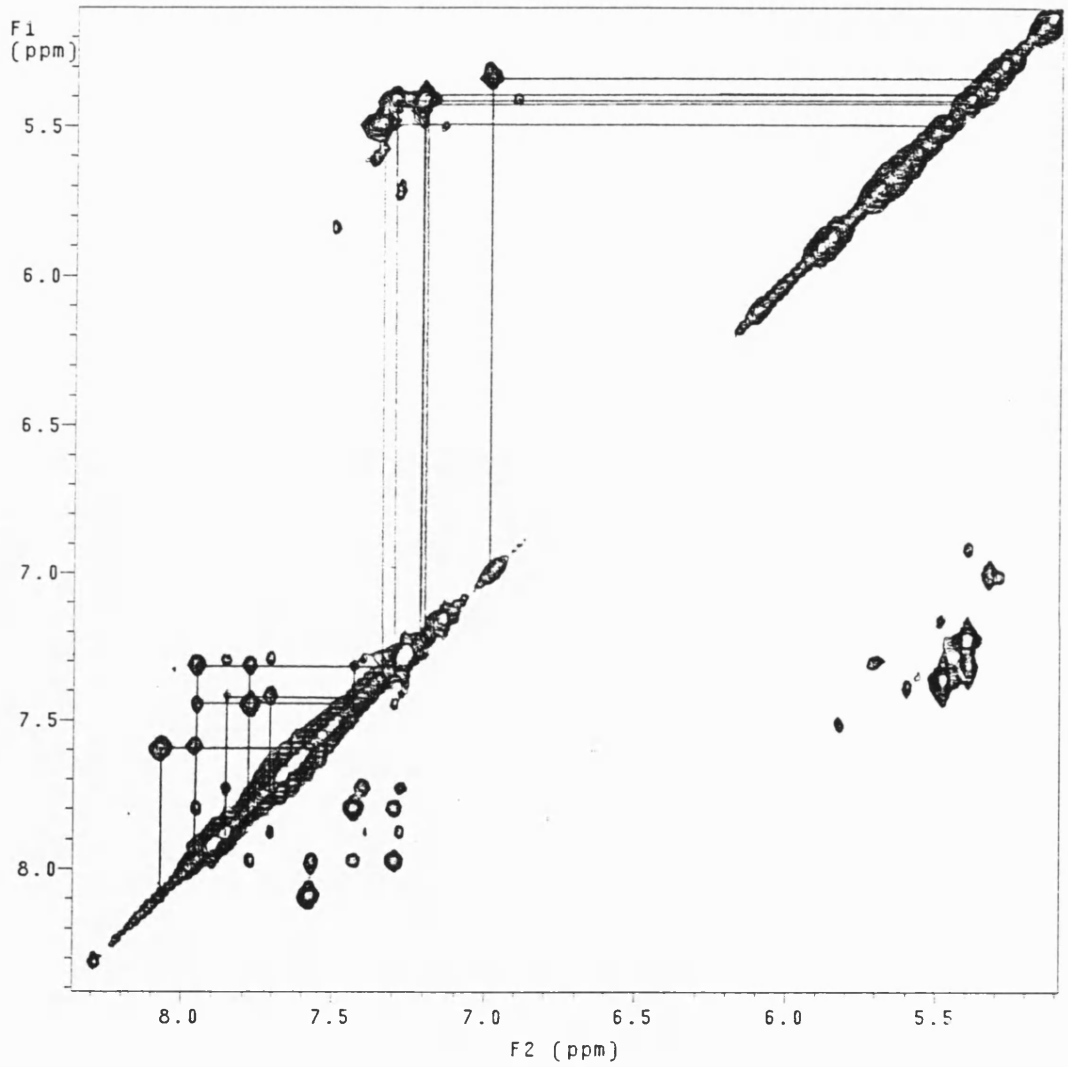


Figure 159 The 2D 600 MHz ^{1}H TOCSY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The aromatics→aromatics/CH5 region

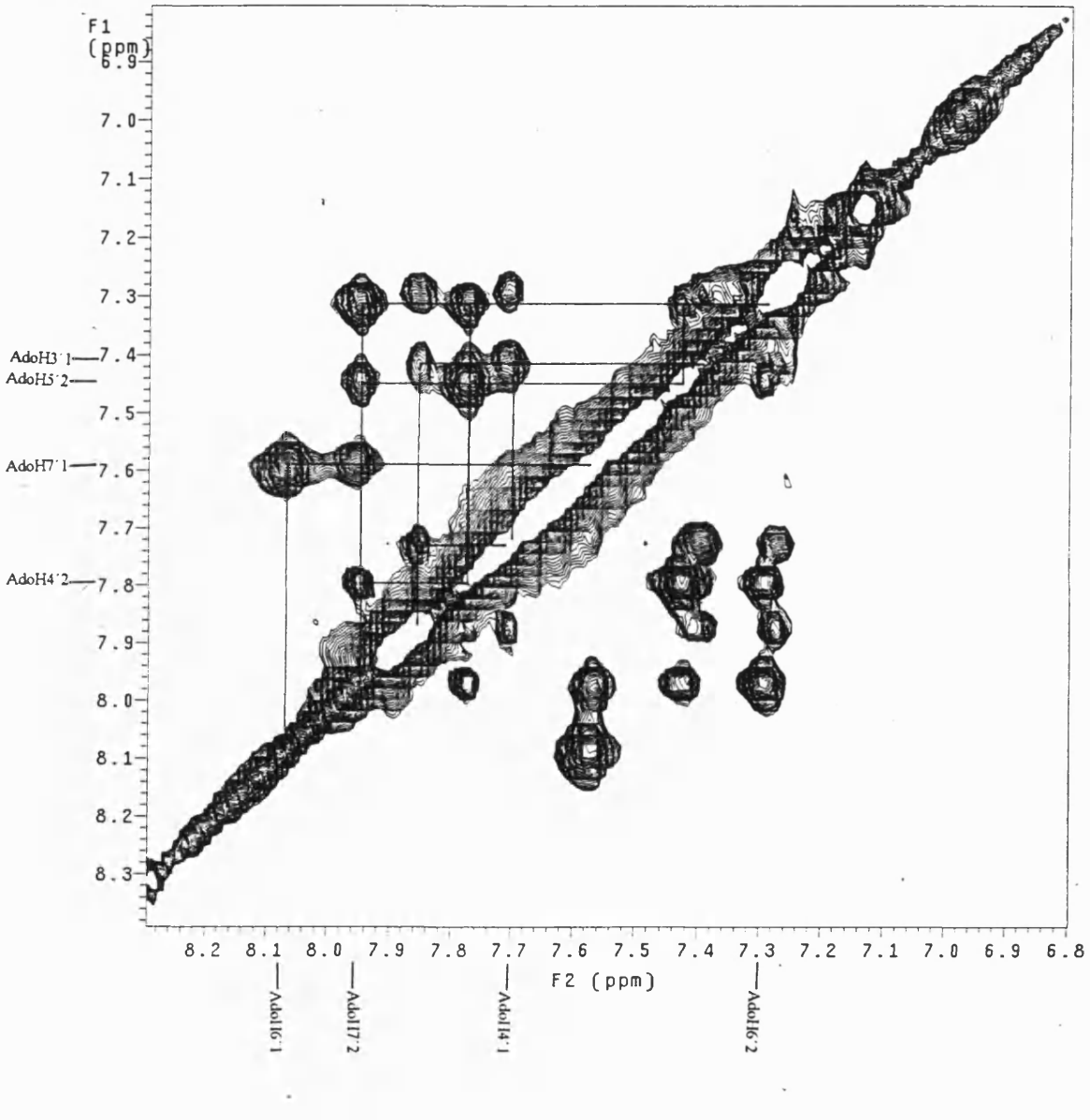


Figure 160 The 2D 600 MHz ^{1}H TOCSY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The aromatics → aromatics region.

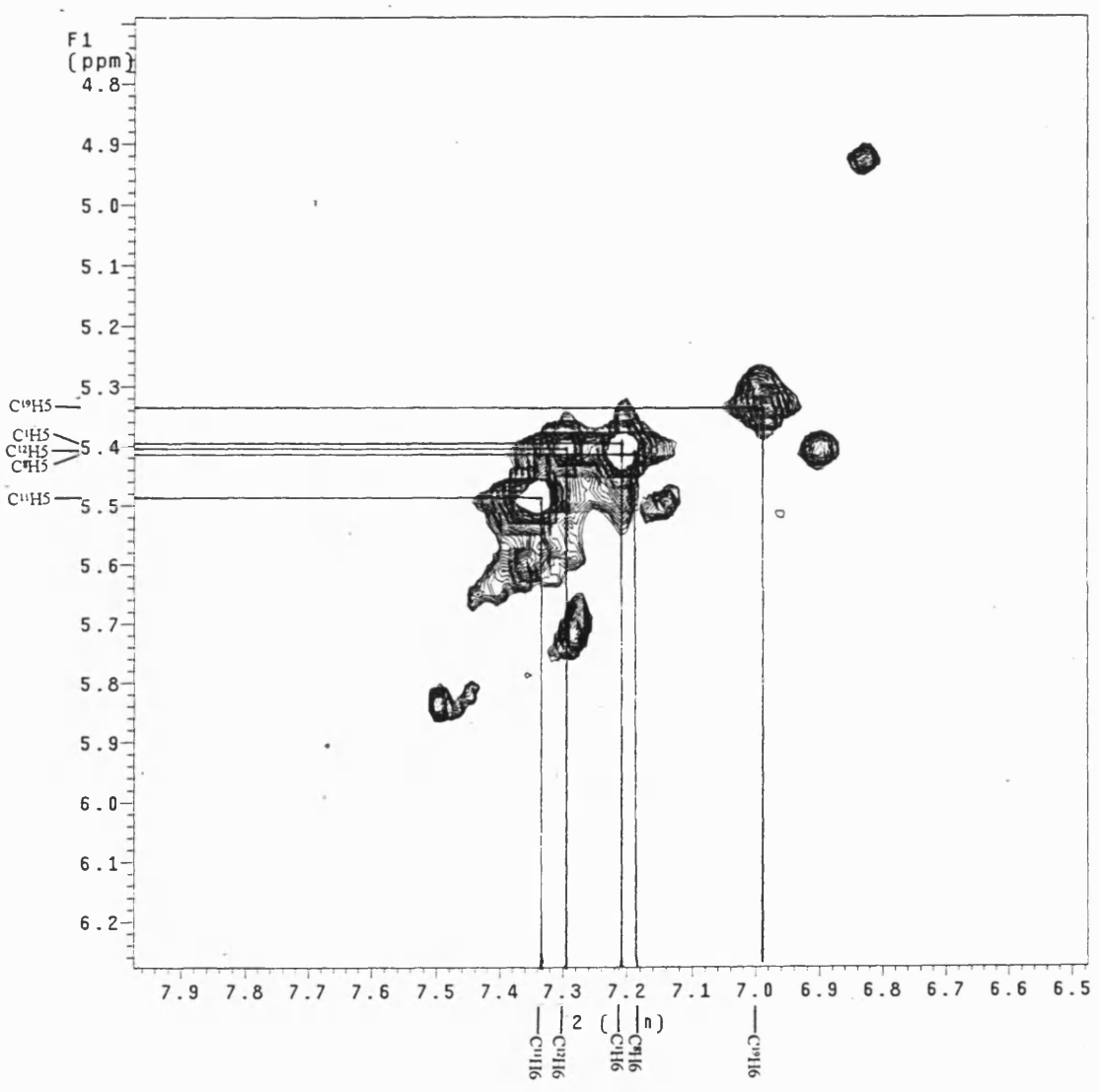


Figure 161 The 2D 600 MHz ^{1}H TOCSY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The aromatics→CH5 region.

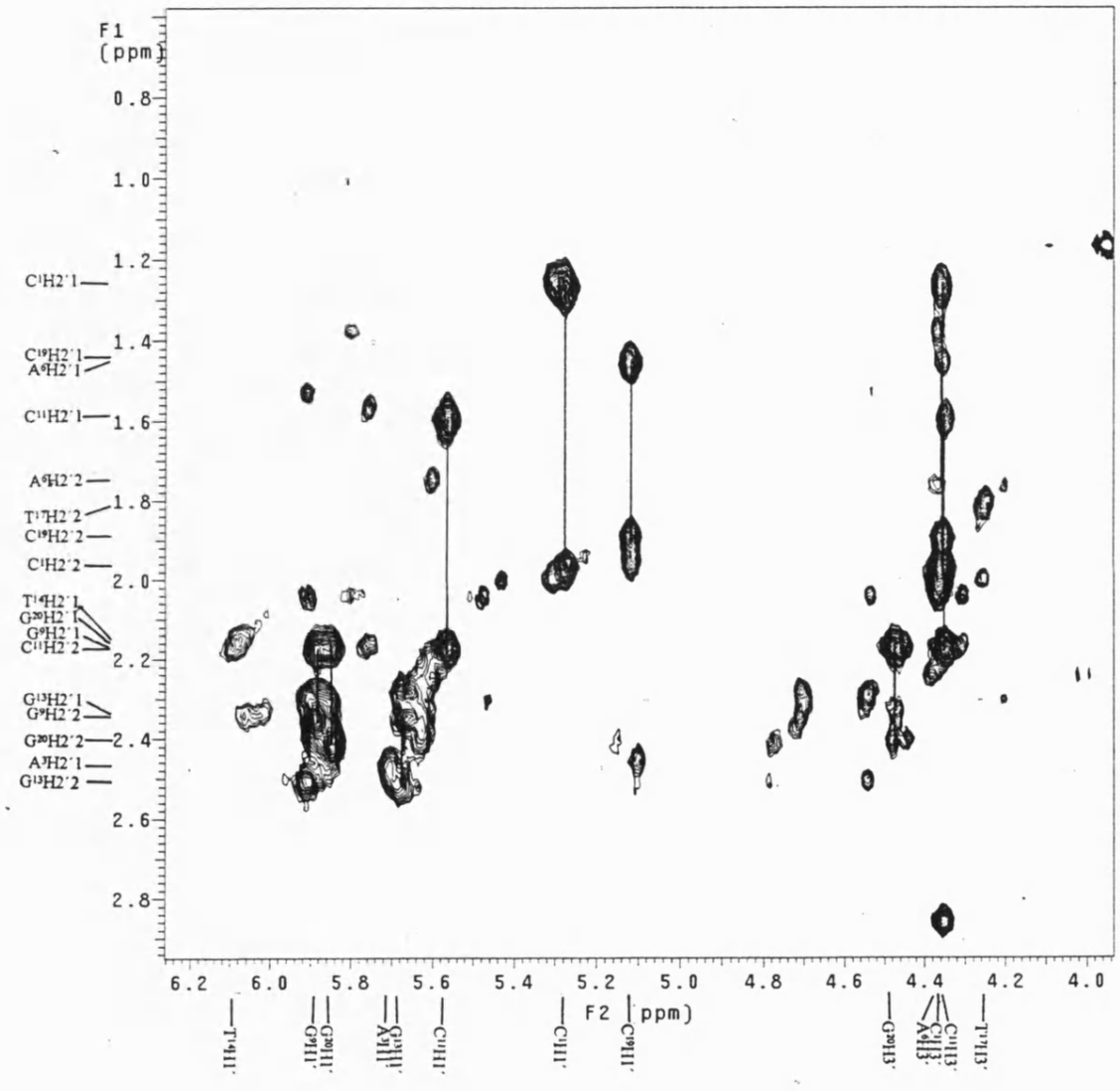


Figure 162 The 2D 600 MHz ^{1}H TOCSY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)-5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The H1′→H2′1/H2′2 region.

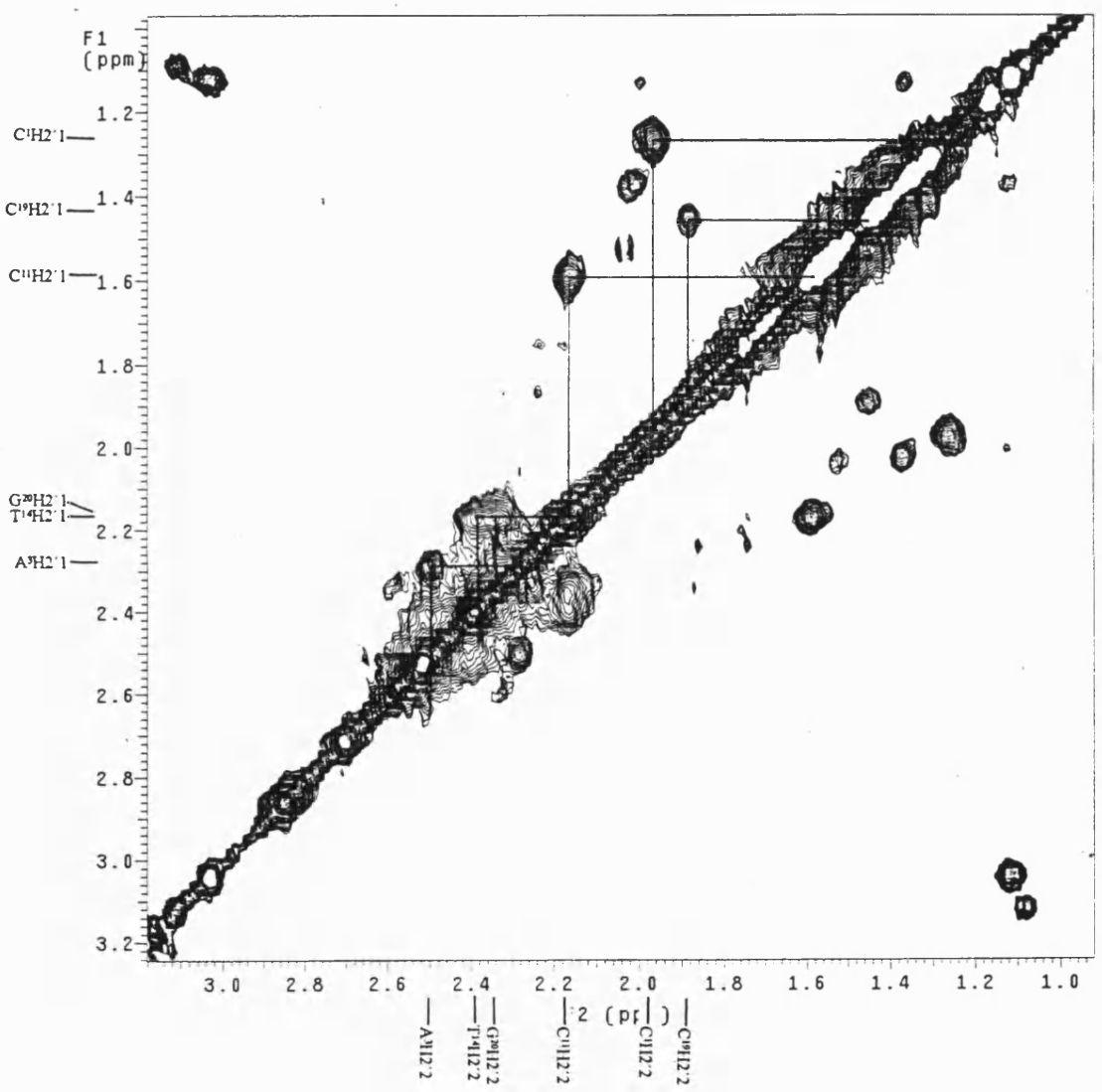


Figure 163 The 2D 600 MHz ^{1}H TOCSY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The H2′1→H2′2 region.

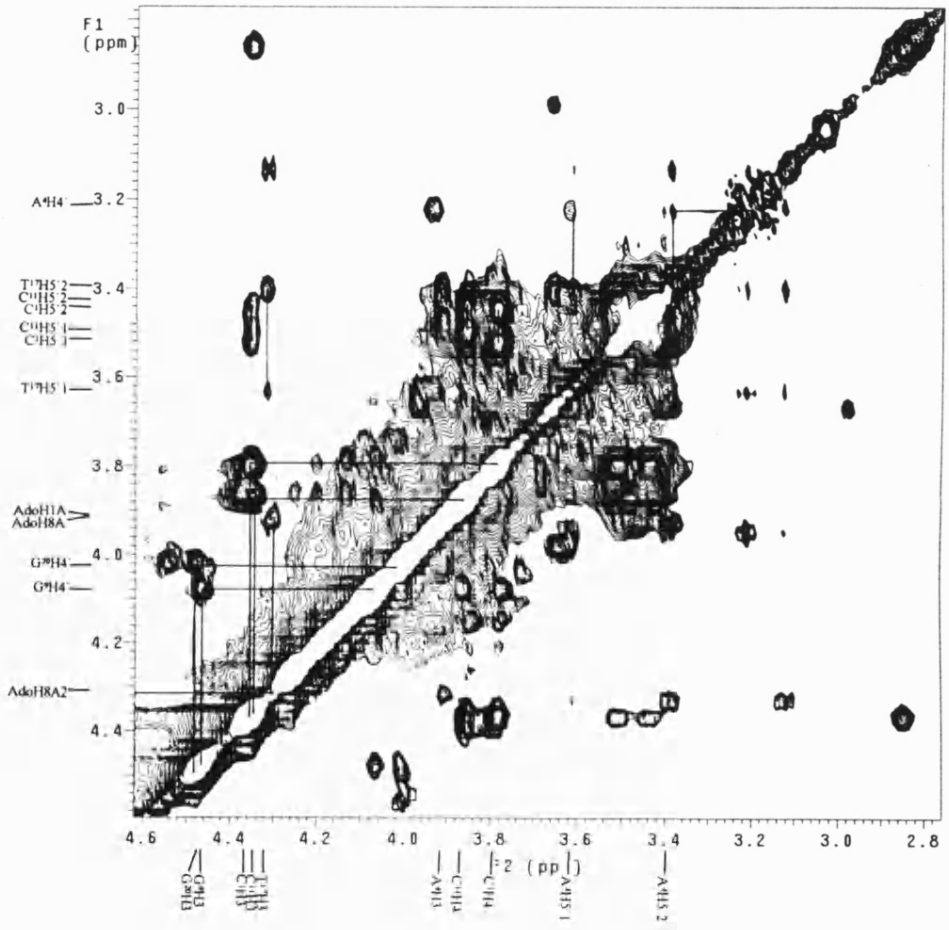


Figure 164 The 2D 600 MHz ^{1}H TOCSY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)-5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. COSY cross-peaks from the H3′, H4′, H5′1 and H5′2 protons.

APPENDIX XVI

The 2D 600 MHz ^{1}H NOESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct.

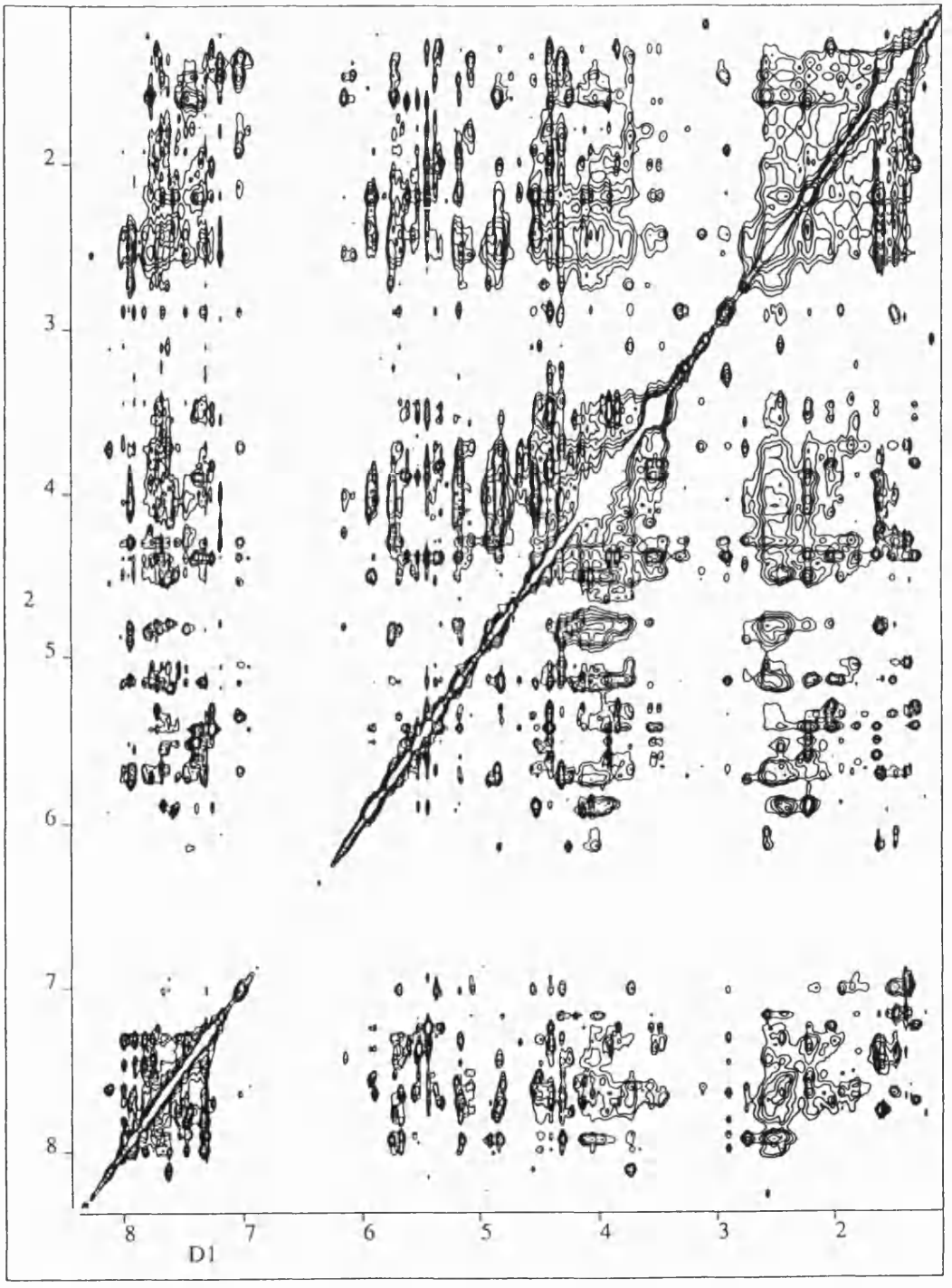


Figure 165 The 2D 600 MHz ^{1}H NOESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct.

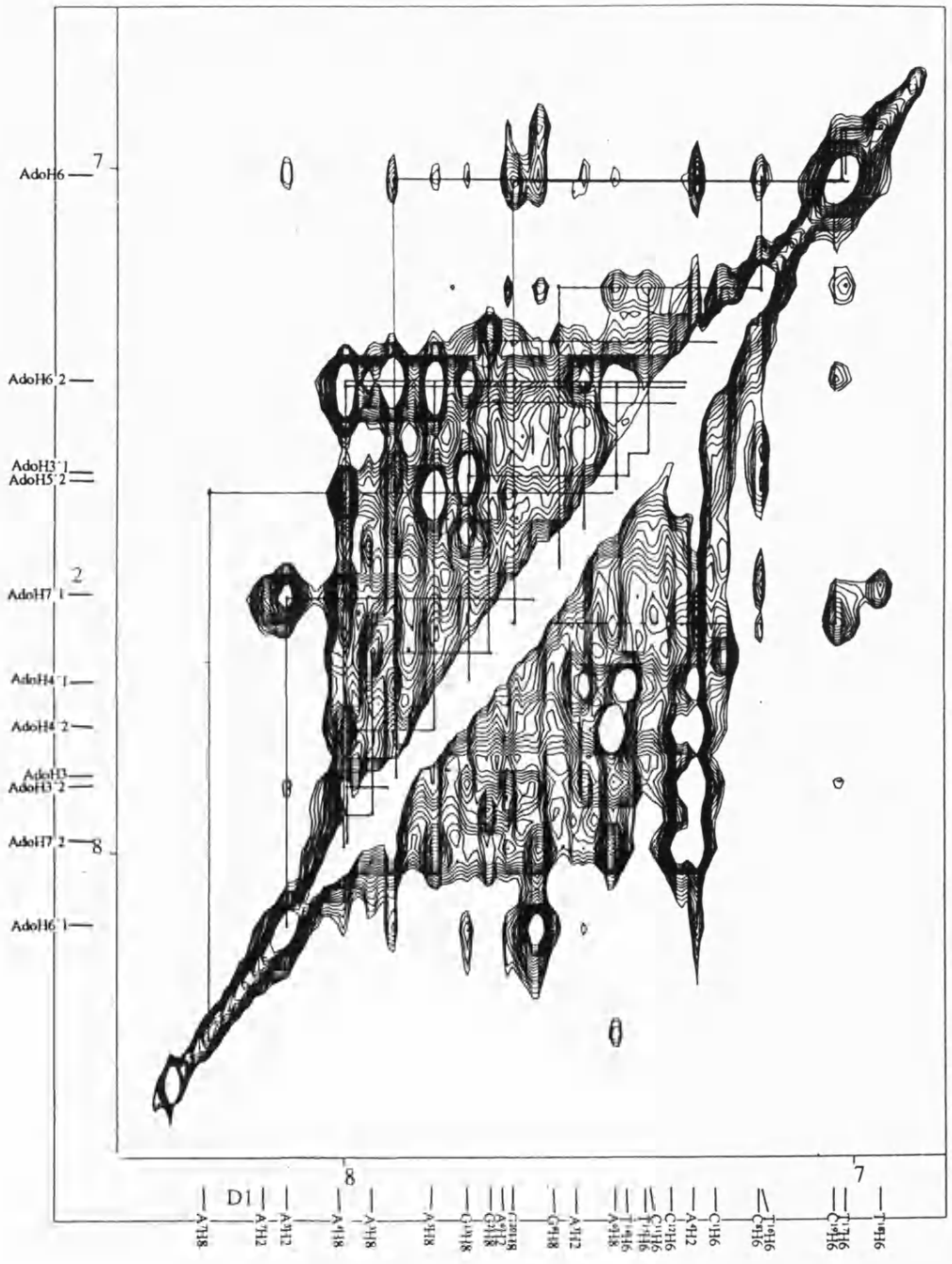


Figure 166 The 2D 600 MHz ^{1}H NOESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The aromatics to aromatics NOE region.

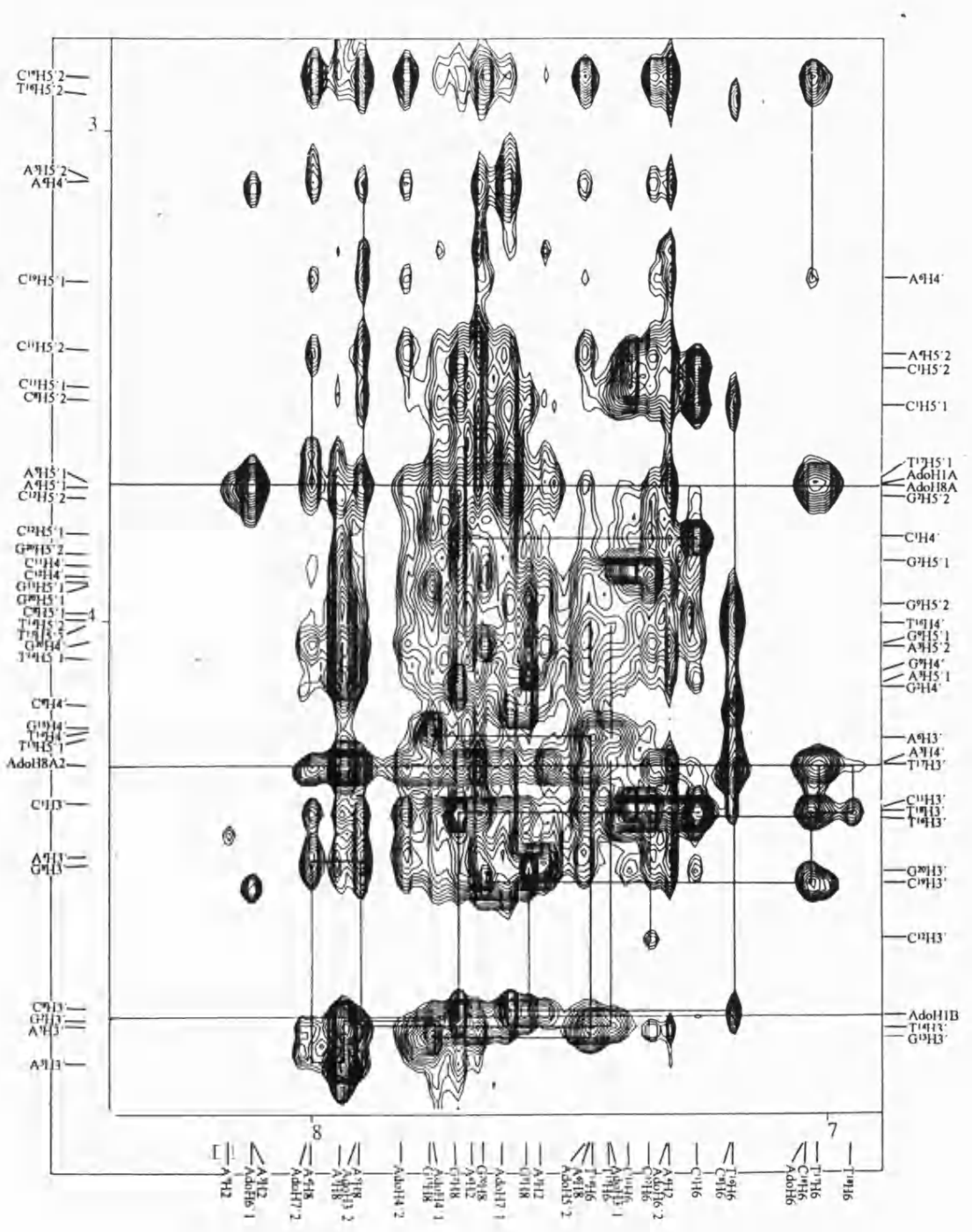


Figure 167 The 2D 600 MHz ^{1}H NOESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7•}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The aromatics→H3′/H4′/H5′1/H5′2 region.

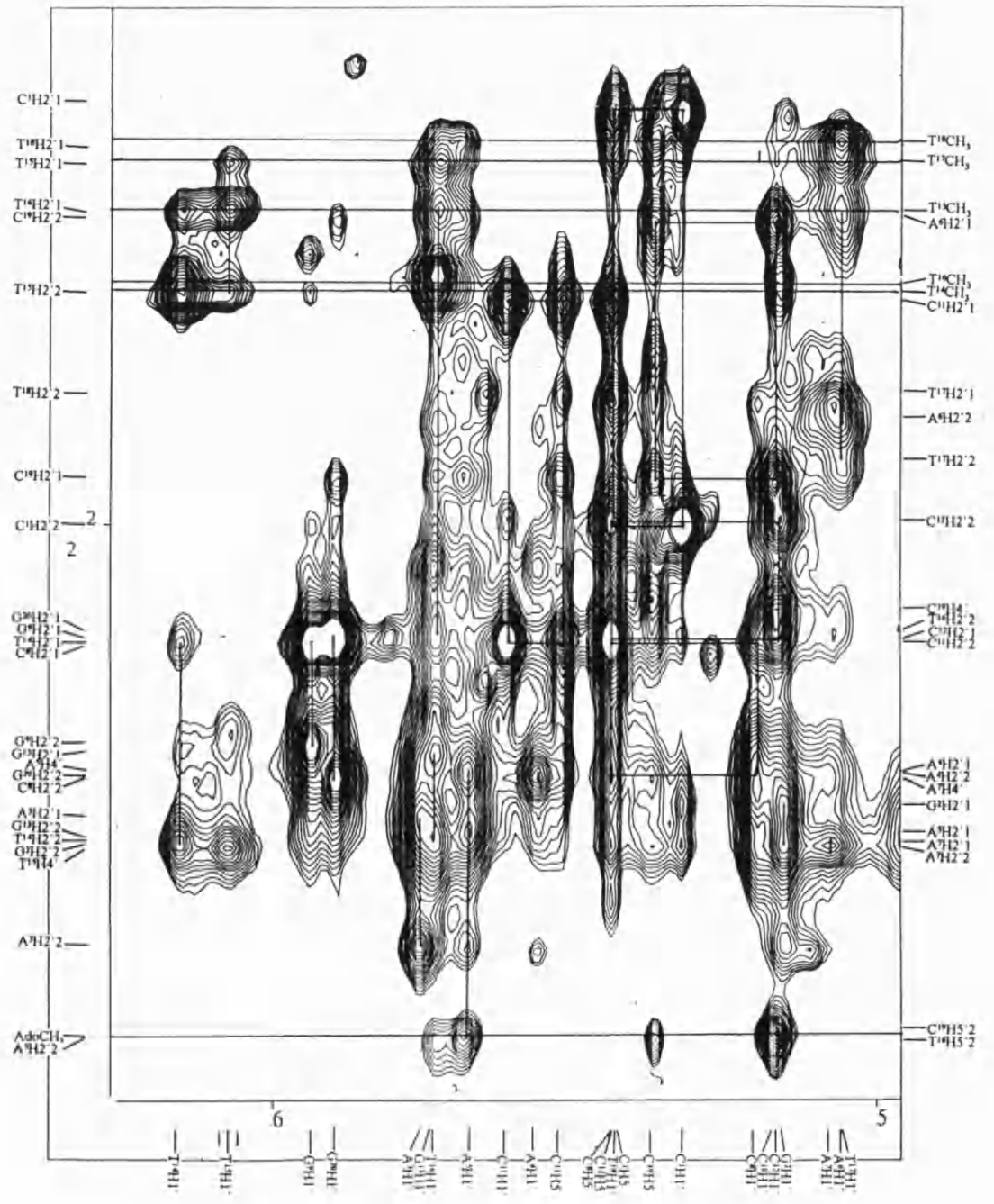


Figure 168 The 2D 600 MHz ^{1}H NOESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The H1′→H2′1/H2′2 region.

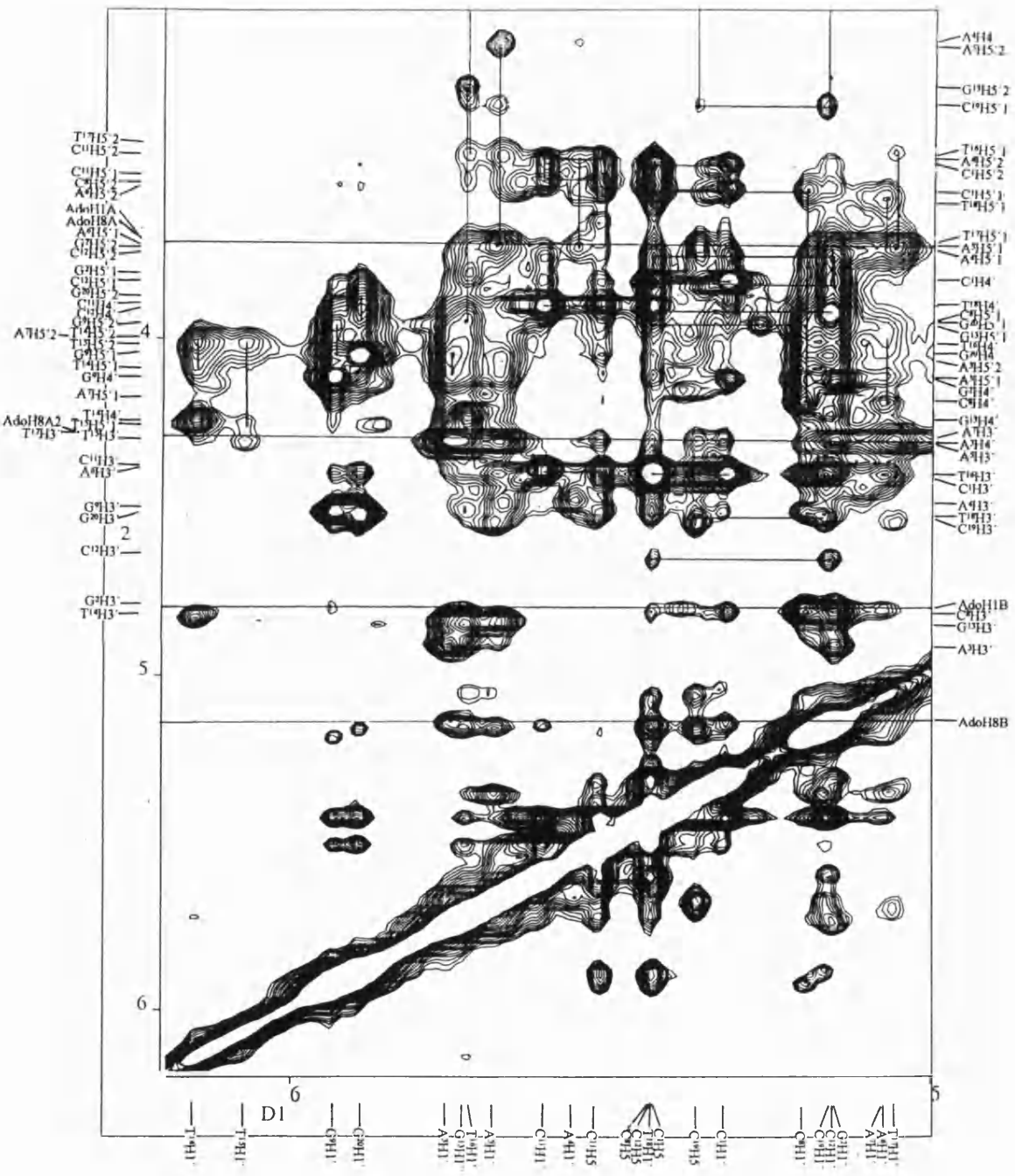


Figure 169 The 2D 600 MHz ^{1}H NOESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The H1′→H3′/H4′/H5′1/H5′2 region.

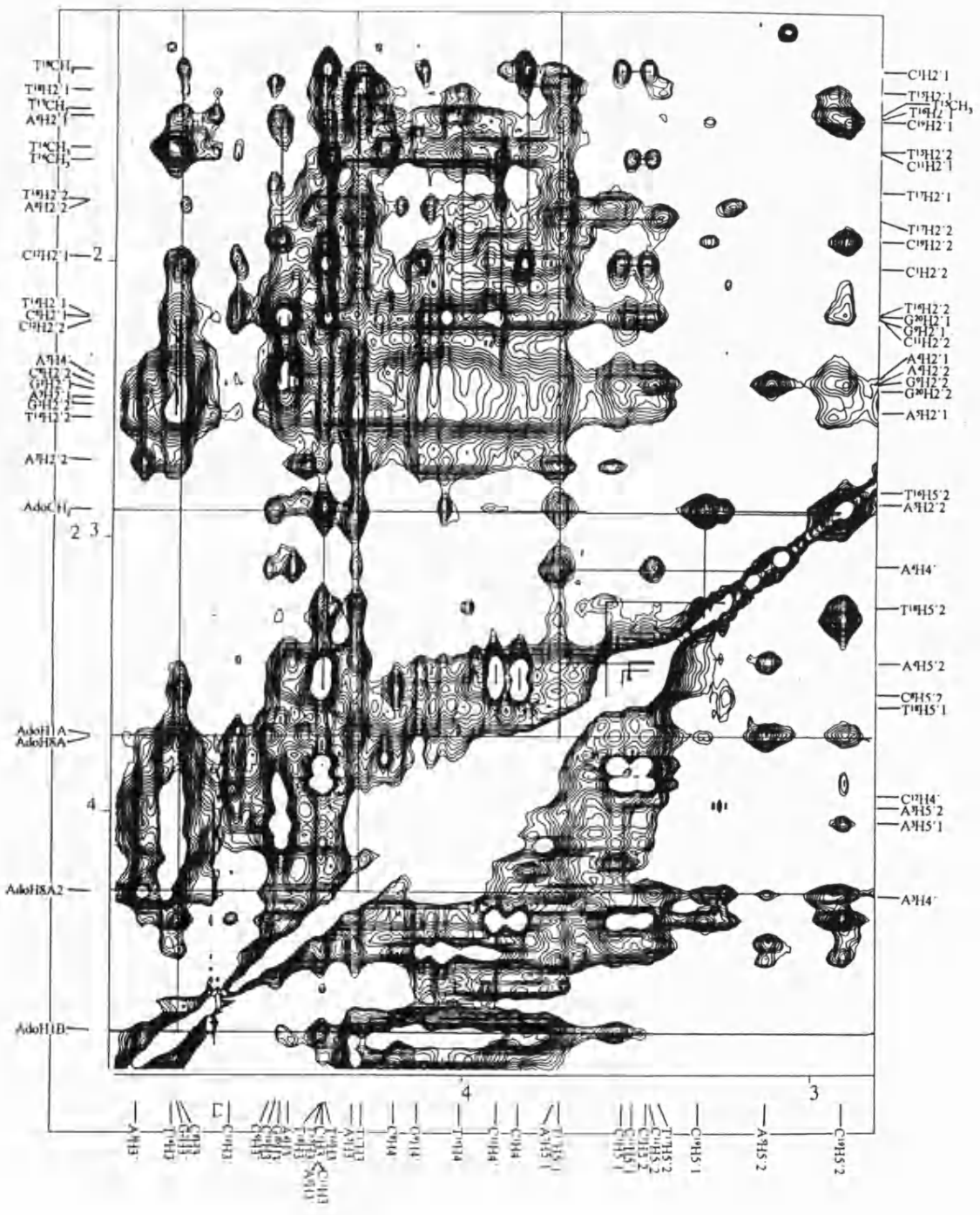


Figure 170 The 2D 600 MHz ^{1}H NOESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7\bullet}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The H3′, H4′, H5′1, H5′2, H2′1 and H2′2 NOE cross-connectivities.

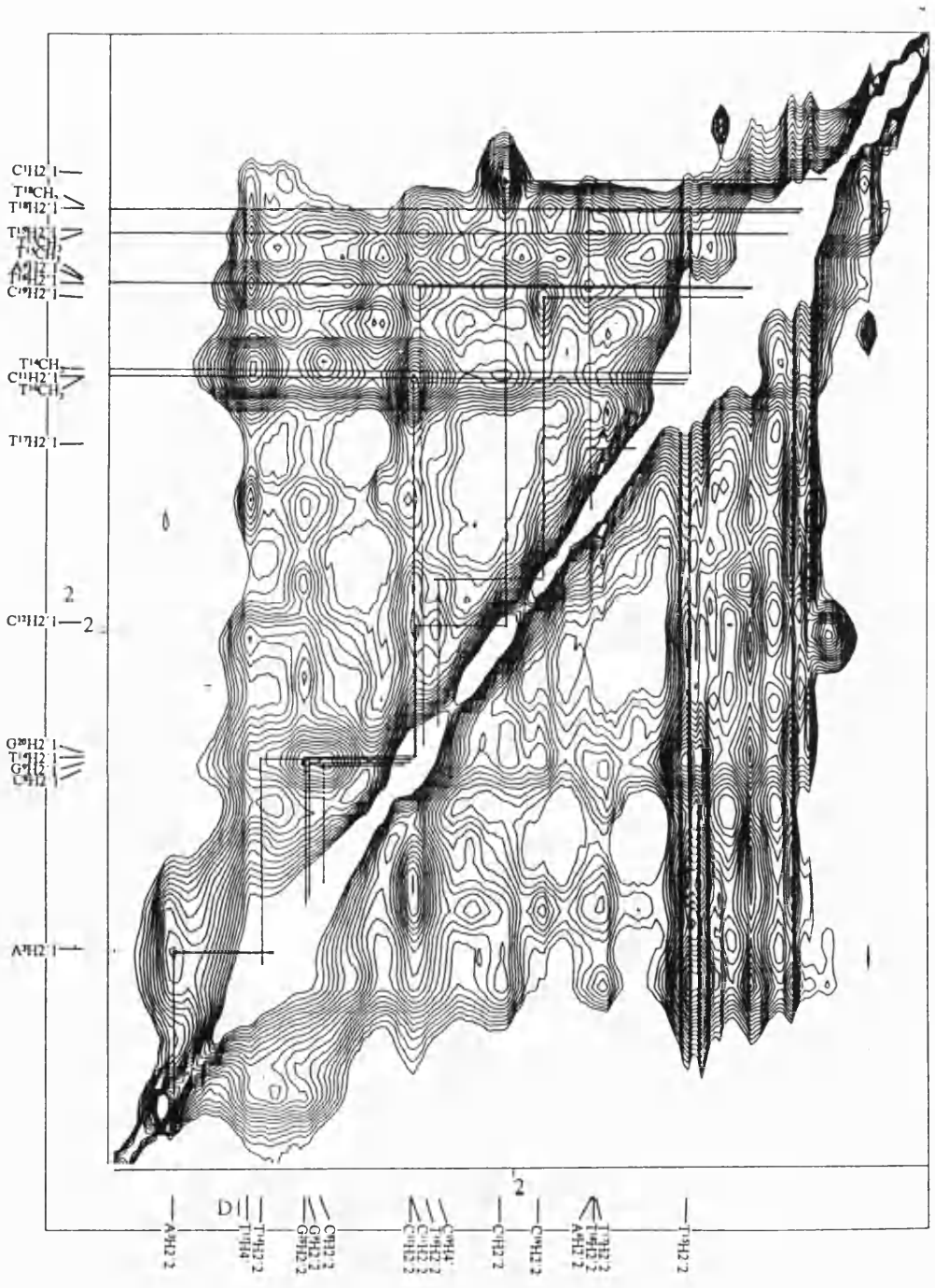


Figure 171 The 2D 600 MHz ^{1}H NOESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7•}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The H2′1→H2′2 region.

APPENDIX XVII

The 2D 600 MHz ^{1}H ROESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct.

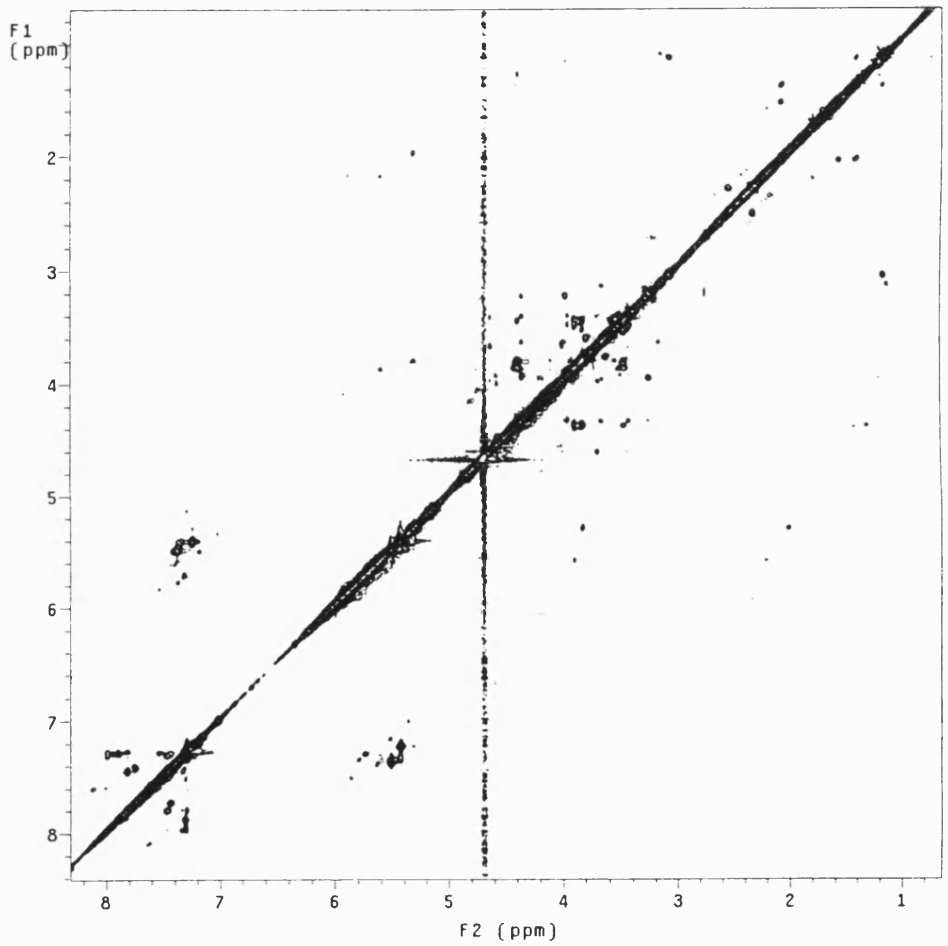


Figure 172 The 2D 600 MHz ^{1}H ROESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct.

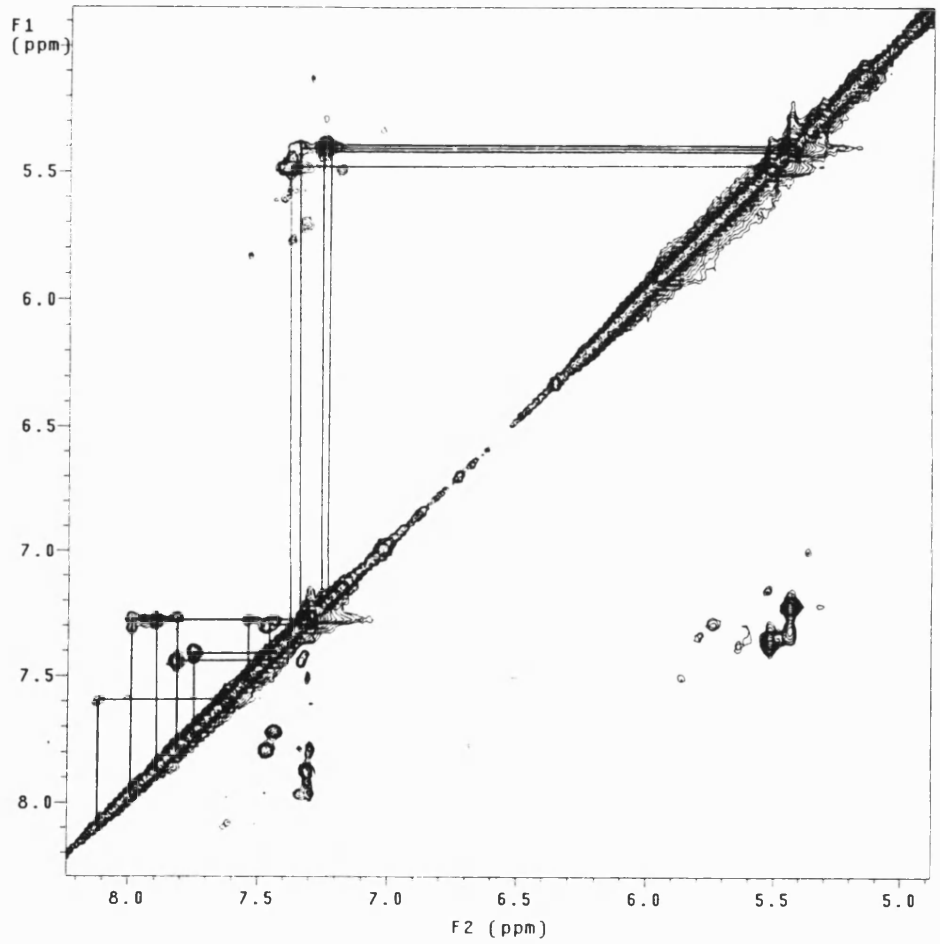


Figure 173 The 2D 600 MHz ^{1}H ROESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The aromatics→aromatics/CH5 region.

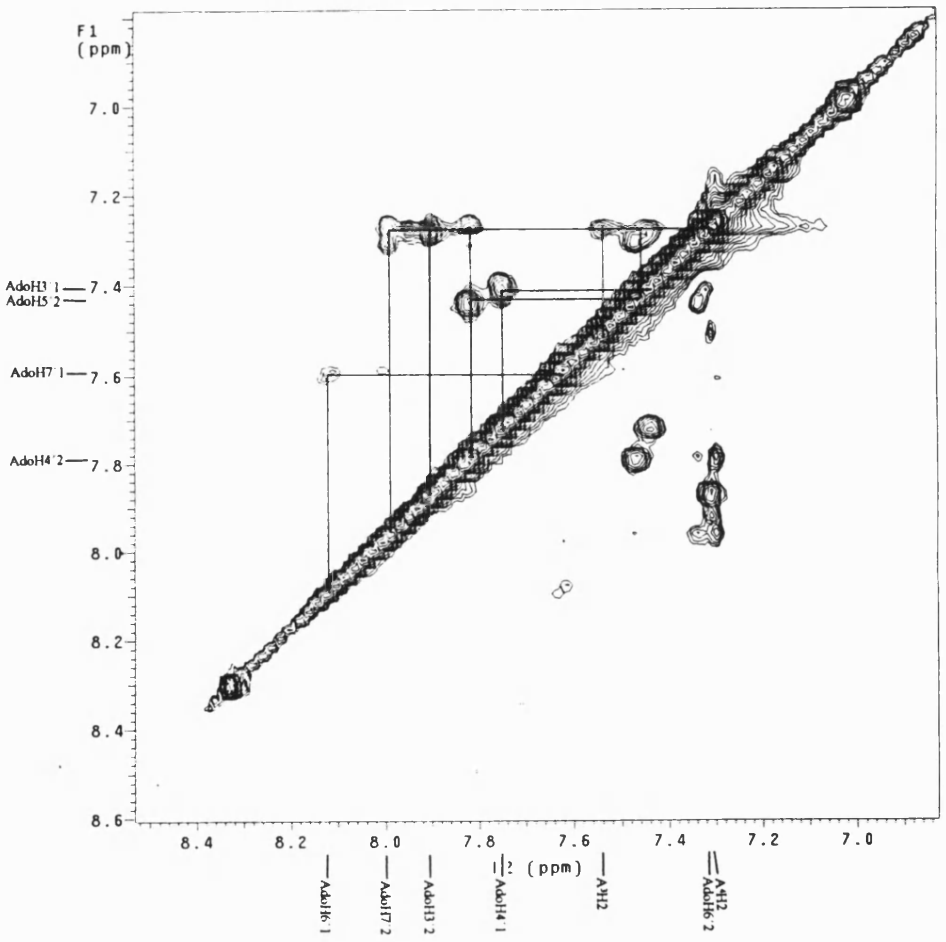


Figure 174 The 2D 600 MHz ^{1}H ROESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The aromatics→aromatics region.

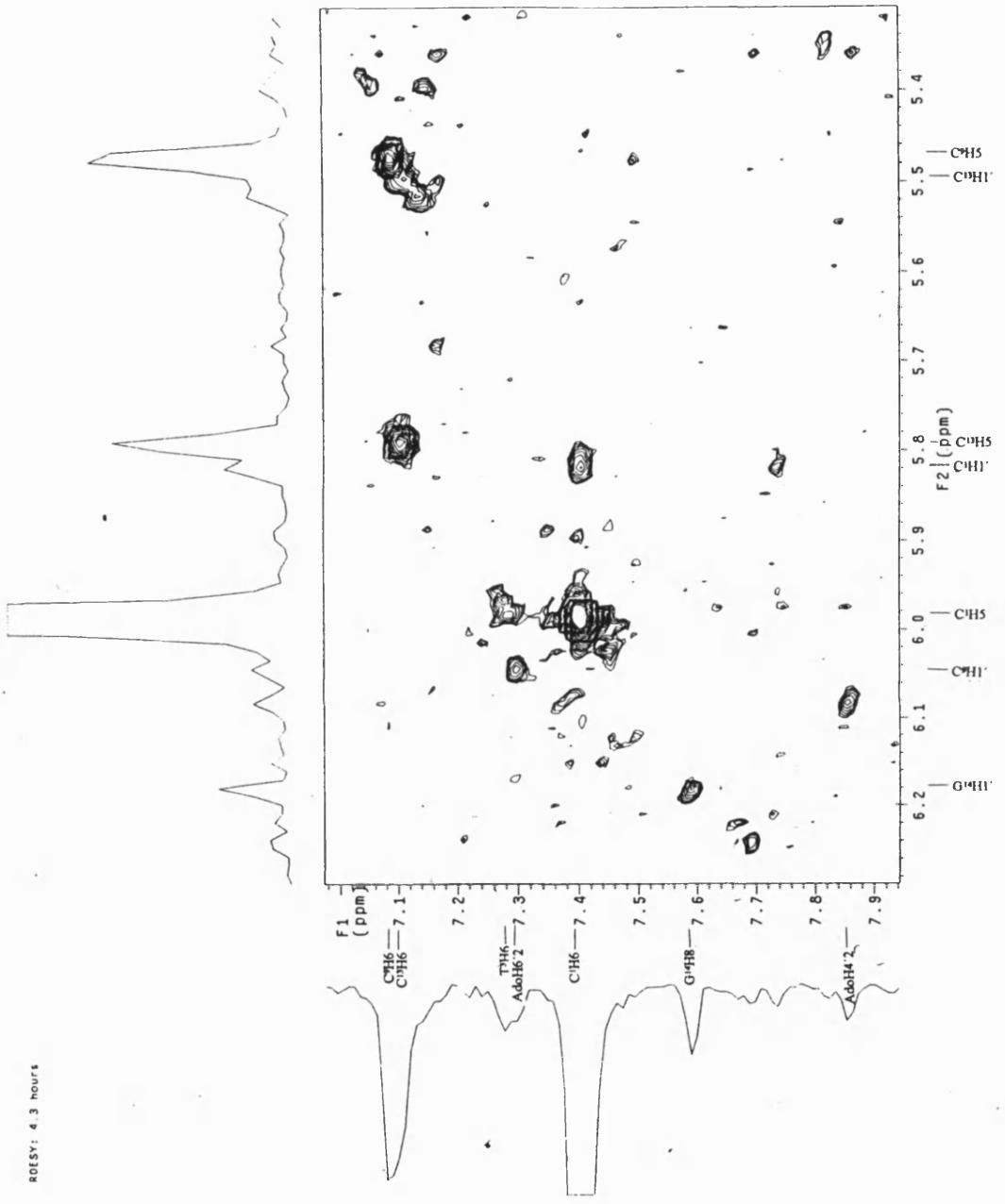


Figure 145 The 600 MHz ROESY ^{1}H NMR spectrum of the adozelesin-5′-($C^{1}G^{2}T^{3}A^{4}A^{5}G^{6}C^{7}G^{8}C^{9}T^{10}T^{11}A^{12•}C^{13}G^{14}$)$_2$ DNA adduct. The aromatics→CH5 region.

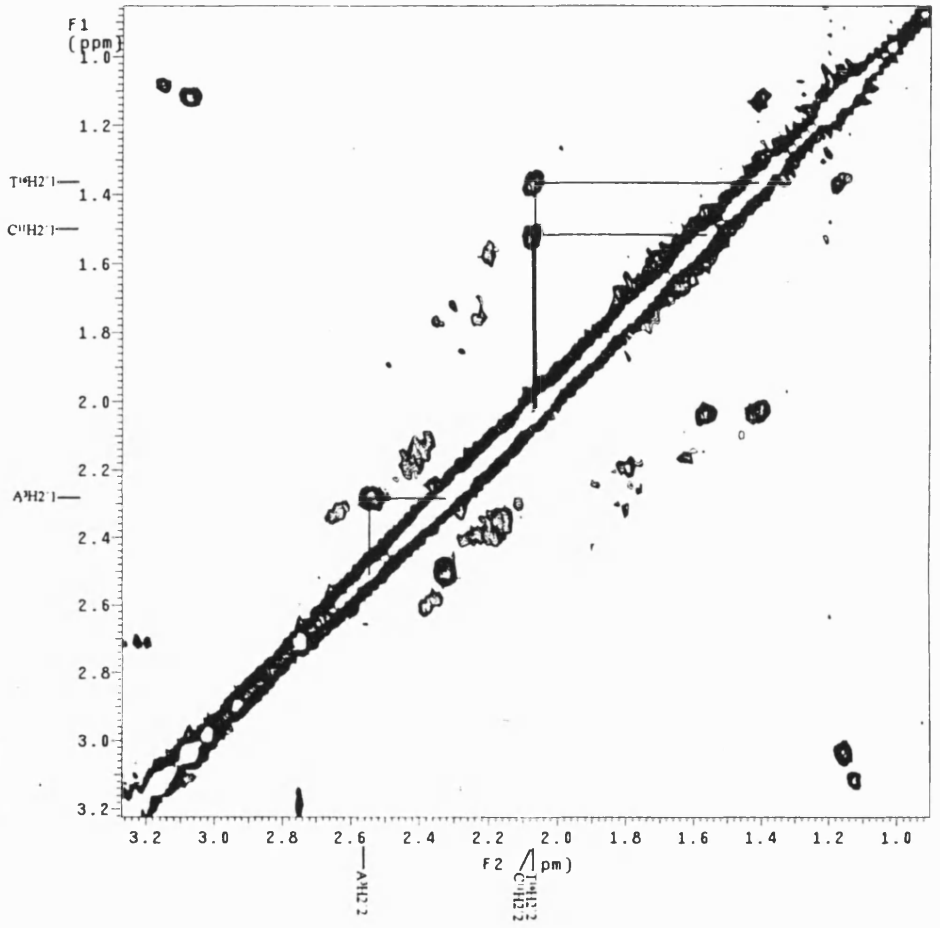


Figure 176 The 2D 600 MHz ^{1}H ROESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The H2′1→H2′2 region.

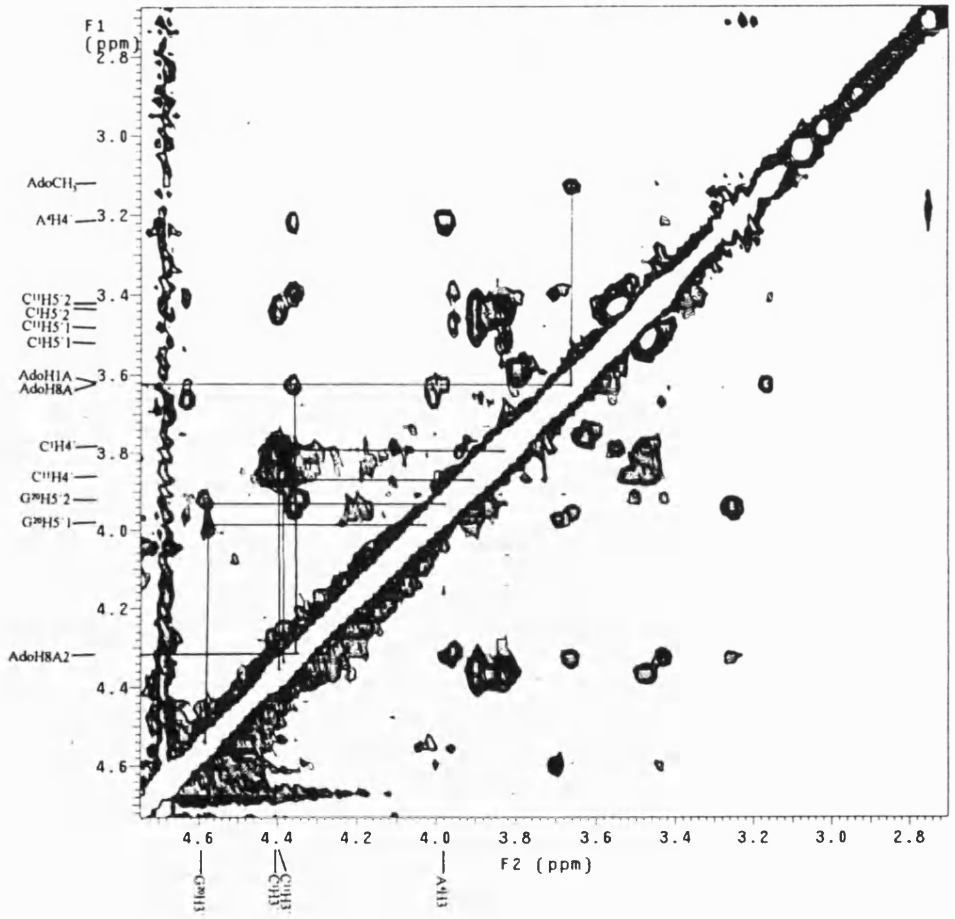


Figure 177 The 2D 600 MHz ^{1}H ROESY spectrum of the adozelesin-5′-($C^{1}G^{2}A^{3}A^{4}A^{5}A^{6}A^{7*}C^{8}G^{9}G^{10}$)-5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct. The cross-connectivities of the H3′, H4′, H5′1 and H5′2 protons.

APPENDIX XVIII

The Peak Pick Spreadsheet for the adozelesin-5′-($C^1G^2A^3A^4A^5A^6A^{7*}C^8G^9G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct.

Table 8 The peak pick spreadsheet.

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
1: P1	7.25	5.32	2.6e+08	A/C1.H6	A/C1.H1'	1.49	5.00
2: P2	7.72	5.32	2.9e+08	A/G2.H8	A/C1.H1'	1.49	3.00
3: P3	7.72	5.14	3.5e+08	A/G2.H8	A/G2.H1'	1.49	3.00
4: P4	7.94	5.14	2.7e+08	A/A3.H8	A/G2.H1'	1.49	5.00
5: P5	7.94	5.75	4.7e+08	A/A3.H8	A/A3.H1'	1.49	3.00
6: P6	8.00	5.75	1.2e+08	A/A4.H8	A/A3.H1'	1.49	7.00
7: P7	8.00	5.55	4.1e+07	A/A4.H8	A/A4.H1'	1.49	7.00
8: P8	7.91	5.55	9.4e+07	A/A5.H8	A/A4.H1'	1.49	7.00
9: P9	7.91	5.67	2.6e+08	A/A5.H8	A/A5.H1'	1.49	5.00
10: P10	7.47	5.67	1.5e+08	A/A6.H8	A/A5.H1'	1.49	5.00
11: P11	7.47	5.07	1.2e+08	A/A6.H8	A/A6.H1'	1.49	7.00
12: P12	8.27	5.07	1.4e+07	A/A7.H8	A/A6.H1'	1.49	7.00
13: P13	8.27	5.08	1.2e+06	A/A7.H8	A/A7.H1'	1.49	7.00
14: P14	7.18	5.08	5.8e+07	A/C8.H6	A/A7.H1'	1.49	7.00
15: P15	7.18	5.19	3.5e+07	A/C8.H6	A/C8.H1'	1.49	7.00
16: P16	7.58	5.20	9.9e+07	A/G9.H8	A/C8.H1'	1.49	7.00
17: P17	7.58	5.92	3.1e+08	A/G9.H8	A/G9.H1'	1.49	3.00
18: P20	7.39	5.61	2.9e+08	B/C11.H6	B/C11.H1'	1.49	3.00
19: P21	7.34	5.61	2e+08	B/C12.H6	B/C11.H1'	1.49	5.00
20: P24	7.77	5.73	2.4e+08	B/G13.H8	B/G13.H1'	1.49	5.00
21: P25	7.46	5.73	1.8e+08	B/T14.H6	B/G13.H1'	1.49	5.00
22: P26	7.46	6.14	1.4e+08	B/T14.H6	B/T14.H1'	1.49	5.00
23: P27	7.41	6.14	8.9e+07	B/T15.H6	B/T14.H1'	1.49	7.00
24: P28	7.41	6.06	4.9e+07	B/T15.H6	B/T15.H1'	1.49	7.00
25: P29	7.18	6.06	4.9e+07	B/T16.H6	B/T15.H1'	1.49	7.00
26: P30	7.18	5.71	3e+07	B/T16.H6	B/T16.H1'	1.49	7.00
27: P31	7.02	5.72	1.1e+08	B/T17.H6	B/T16.H1'	1.49	7.00
28: P32	7.02	5.06	1.5e+08	B/T17.H6	B/T17.H1'	1.49	5.00
29: P33	6.95	5.06	9.7e+07	B/T18.H6	B/T17.H1'	1.49	7.00
30: P34	6.95	5.44	6.8e+07	B/T18.H6	B/T18.H1'	1.49	7.00
31: P35	7.03	5.16	1e+08	B/C19.H6	B/C19.H1'	1.49	7.00
32: P36	7.66	5.16	2.4e+08	B/G20.H8	B/C19.H1'	1.49	5.00
33: P37	7.67	5.90	2.6e+08	B/G20.H8	B/G20.H1'	1.49	5.00
34: P38	7.26	1.31	3.7e+08	A/C1.H6	A/C1.H2'1	1.49	3.00
35: P39	7.25	2.01	4.3e+08	A/C1.H6	A/C1.H2'2	1.49	3.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
36: P40	7.72	1.31	2.2e+08	A/G2.H8	A/C1.H2'1	1.49	5.00
37: P41	7.72	2.01	2.7e+08	A/G2.H8	A/C1.H2'2	1.49	5.00
38: P42	7.72	2.50	2.7e+08	A/G2.H8	A/G2.H2'1	1.49	3.00
39: P43	7.72	2.56	3e+08	A/G2.H8	A/G2.H2'2	1.49	3.00
40: P44	7.94	2.50	1.5e+08	A/A3.H8	A/G2.H2'1	1.49	5.00
41: P45	7.94	2.56	1.6e+08	A/A3.H8	A/G2.H2'2	1.49	5.00
42: P46	7.94	2.73	4.8e+08	A/A3.H8	A/A3.H2'2	1.49	3.00
43: P47	7.94	2.52	8.3e+08	A/A3.H8	A/A3.H2'1	1.49	3.00
44: P48	8.00	2.73	8.8e+07	A/A4.H8	A/A3.H2'2	1.49	7.00
45: P49	8.00	2.45	2.3e+08	A/A4.H8	A/A4.H2'2	1.49	5.00
46: P50	8.00	2.52	1.6e+08	A/A4.H8	A/A3.H2'1	1.49	5.00
47: P51	8.00	2.43	2.3e+08	A/A4.H8	A/A4.H2'1	1.49	5.00
48: P52	7.90	2.54	1.9e+08	A/A5.H8	A/A5.H2'1	1.49	5.00
49: P53	7.90	2.90	1.8e+08	A/A5.H8	A/A5.H2'2	1.49	5.00
50: P54	7.90	2.43	1.6e+08	A/A5.H8	A/A4.H2'1	1.49	5.00
51: P55	7.90	2.45	1.7e+08	A/A5.H8	A/A4.H2'2	1.49	5.00
52: P56	7.47	2.90	4.5e+07	A/A6.H8	A/A5.H2'2	1.49	7.00
53: P57	7.47	2.54	1.9e+08	A/A6.H8	A/A5.H2'1	1.49	5.00
54: P58	7.47	1.47	2.3e+08	A/A6.H8	A/A6.H2'1	1.49	5.00
55: P59	7.47	1.81	1.4e+08	A/A6.H8	A/A6.H2'2	1.49	7.00
56: P60	8.26	2.56	4e+07	A/A7.H8	A/A7.H2'1	1.49	7.00
57: P61	8.26	2.57	4.4e+07	A/A7.H8	A/A7.H2'2	1.49	7.00
58: P62	7.18	2.56	1.5e+08	A/C8.H6	A/A7.H2'1	1.49	5.00
59: P63	7.18	2.57	1.7e+08	A/C8.H6	A/A7.H2'2	1.49	5.00
60: P64	7.58	2.21	1.8e+08	A/G9.H8	A/C8.H2'1	1.49	5.00
61: P65	7.18	2.21	5.4e+07	A/C8.H6	A/C8.H2'1	1.49	7.00
62: P66	7.18	2.46	2.9e+08	A/C8.H6	A/C8.H2'2	1.49	3.00
63: P67	7.25	2.21	1.8e+08	A/G10.H8	A/G9.H2'2	1.49	5.00
64: P68	7.58	2.39	3.2e+08	A/G9.H8	A/G9.H2'1	1.49	3.00
65: P69	7.58	2.46	2.4e+08	A/G9.H8	A/C8.H2'2	1.49	5.00
66: P70	7.58	2.21	1.8e+08	A/G9.H8	A/G9.H2'2	1.49	5.00
67: P71	7.25	2.39	2.2e+08	A/G10.H8	A/G9.H2'1	1.49	5.00
68: P74	7.39	2.21	2.3e+08	B/C11.H6	B/C11.H2'2	1.49	5.00
69: P75	7.39	1.62	3.2e+08	B/C11.H6	B/C11.H2'1	1.49	3.00
70: P76	7.35	1.62	4e+08	B/C12.H6	B/C11.H2'1	1.49	3.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
71: P77	7.35	2.21	3.5e+08	B/C12.H6	B/C11.H2'2	1.49	3.00
72: P78	7.35	2.20	2.6e+08	B/C12.H6	B/C12.H2'1	1.49	5.00
73: P79	7.35	1.98	2.2e+08	B/C12.H6	B/C12.H2'2	1.49	5.00
74: P80	7.77	1.99	1.7e+08	B/G13.H8	B/C12.H2'2	1.49	5.00
75: P81	7.77	2.20	3e+08	B/G13.H8	B/C12.H2'1	1.49	3.00
76: P82	7.77	2.40	3.4e+08	B/G13.H8	B/G13.H2'1	1.49	3.00
77: P83	7.77	2.55	5.4e+08	B/G13.H8	B/G13.H2'2	1.49	3.00
78: P84	7.45	2.40	2.1e+08	B/T14.H6	B/G13.H2'1	1.49	5.00
79: P85	7.45	2.21	3.4e+08	B/T14.H6	B/T14.H2'1	1.49	3.00
80: P86	7.46	2.55	1.3e+08	B/T14.H6	B/T14.H2'2	1.49	7.00
81: P87	7.45	2.55	1.3e+08	B/T14.H6	B/G13.H2'2	1.49	7.00
82: P88	7.41	1.38	1.4e+08	B/T15.H6	B/T15.H2'1	1.49	5.00
83: P89	7.41	1.61	2.6e+08	B/T15.H6	B/T15.H2'2	1.49	5.00
84: P90	7.41	2.21	2.1e+08	B/T15.H6	B/T14.H2'1	1.49	5.00
85: P91	7.41	2.55	1.2e+08	B/T15.H6	B/T14.H2'2	1.49	7.00
86: P92	7.18	1.38	2.6e+08	B/T16.H6	B/T15.H2'1	1.49	5.00
87: P93	7.18	1.46	5.1e+08	B/T16.H6	B/T16.H2'2	1.49	3.00
88: P94	7.18	1.61	1.2e+08	B/T16.H6	B/T15.H2'2	1.49	7.00
89: P95	7.18	2.19	6.5e+07	B/T16.H6	B/T16.H2'1	1.49	7.00
90: P96	7.02	1.46	2.1e+08	B/T17.H6	B/T16.H2'2	1.49	5.00
91: P97	7.02	2.19	2.2e+08	B/T17.H6	B/T16.H2'1	1.49	5.00
92: P98	7.02	1.72	2.5e+08	B/T17.H6	B/T17.H2'1	1.49	5.00
93: P99	7.02	1.84	3.6e+08	B/T17.H6	B/T17.H2'2	1.49	3.00
94: P100	6.95	1.84	9.7e+07	B/T18.H6	B/T17.H2'2	1.49	7.00
95: P101	6.95	1.73	5.6e+07	B/T18.H6	B/T17.H2'1	1.49	7.00
96: P102	6.95	1.80	1.9e+08	B/T18.H6	B/T18.H2'1	1.49	5.00
97: P104	7.03	1.92	2.7e+08	B/C19.H6	B/C19.H2'1	1.49	5.00
98: P105	7.03	1.49	5.4e+08	B/C19.H6	B/C19.H2'2	1.49	3.00
99: P107	7.03	1.80	1.2e+08	B/C19.H6	B/T18.H2'1	1.49	7.00
100: P108	7.66	1.49	2.4e+08	B/G20.H8	B/C19.H2'2	1.49	5.00
101: P109	7.66	1.93	2.5e+07	B/G20.H8	B/C19.H2'1	1.49	7.00
102: P110	7.66	2.20	3.9e+08	B/G20.H8	B/G20.H2'2	1.49	3.00
103: P111	7.67	2.43	6e+08	B/G20.H8	B/G20.H2'1	1.49	3.00
104: P112	7.39	4.40	2.5e+08	B/C11.H6	B/C11.H3'	1.49	5.00
105: P113	7.34	4.40	2.5e+08	B/C12.H6	B/C11.H3'	1.49	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
106: P114	7.34	4.65	4.9e+08	B/C12.H6	B/C12.H3'	1.49	3.00
107: P115	7.76	4.65	1.4e+07	B/G13.H8	B/C12.H3'	1.49	7.00
108: P116	7.77	4.85	2.4e+08	B/G13.H8	B/G13.H3'	1.49	5.00
109: P117	7.46	4.85	7.4e+07	B/T14.H6	B/G13.H3'	1.49	7.00
110: P118	7.46	4.83	1.4e+08	B/T14.H6	B/T14.H3'	1.49	7.00
111: P119	7.42	4.82	9.6e+07	B/T15.H6	B/T14.H3'	1.49	7.00
112: P120	7.18	4.30	1.6e+08	B/T16.H6	B/T15.H3'	1.49	5.00
113: P121	7.18	4.39	6e+07	B/T16.H6	B/T16.H3'	1.49	7.00
114: P122	7.02	4.39	8.4e+07	B/T17.H6	B/T16.H3'	1.49	7.00
115: P123	6.95	4.40	7.2e+07	B/T18.H6	B/T18.H3'	1.49	7.00
116: P124	7.02	4.30	2.2e+08	B/T17.H6	B/T17.H3'	1.49	5.00
117: P125	6.95	4.30	7.2e+07	B/T18.H6	B/T17.H3'	1.49	7.00
118: P126	7.03	4.40	6.9e+07	B/C19.H6	B/T18.H3'	1.49	7.00
119: P127	7.03	4.54	1.3e+08	B/C19.H6	B/C19.H3'	1.49	7.00
120: P128	7.66	4.54	2.3e+08	B/G20.H8	B/C19.H3'	1.49	5.00
121: P129	7.25	4.39	2.4e+08	A/C1.H6	A/C1.H3'	1.49	5.00
122: P130	7.71	4.39	2.6e+08	A/G2.H8	A/C1.H3'	1.49	5.00
123: P131	7.72	4.81	2.2e+09	A/G2.H8	A/G2.H3'	1.49	3.00
124: P132	7.94	4.81	2.6e+08	A/A3.H8	A/G2.H3'	1.49	5.00
125: P133	7.94	4.89	4.3e+08	A/A3.H8	A/A3.H3'	1.49	3.00
126: P134	8.00	4.89	6.4e+07	A/A4.H8	A/A3.H3'	1.49	7.00
127: P135	8.00	4.50	1.3e+08	A/A4.H8	A/A4.H3'	1.49	7.00
128: P136	7.90	4.50	1.7e+08	A/A5.H8	A/A4.H3'	1.49	5.00
129: P137	7.90	4.30	1.1e+08	A/A5.H8	A/A5.H3'	1.49	7.00
130: P138	7.47	4.30	2e+08	A/A6.H8	A/A5.H3'	1.49	5.00
131: P139	7.47	4.40	1.2e+08	A/A6.H8	A/A6.H3'	1.49	7.00
132: P140	7.19	4.30	1e+08	A/C8.H6	A/A7.H3'	1.49	7.00
133: P141	7.18	4.80	8.5e+07	A/C8.H6	A/C8.H3'	1.49	7.00
134: P142	7.58	4.80	1.3e+08	A/G9.H8	A/C8.H3'	1.49	7.00
135: P143	7.58	4.51	2.4e+08	A/G9.H8	A/G9.H3'	1.49	5.00
136: P144	7.25	4.51	5.5e+07	A/G10.H8	A/G9.H3'	1.49	7.00
137: P146	7.66	4.52	1.4e+08	B/G20.H8	B/G20.H3'	1.49	7.00
138: P147	7.26	3.83	1.8e+08	A/C1.H6	A/C1.H4'	1.49	5.00
139: P148	7.71	3.83	2.5e+08	A/G2.H8	A/C1.H4'	1.49	5.00
140: P149	7.72	4.12	1.8e+08	A/G2.H8	A/G2.H4'	1.49	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
141: P150	7.94	4.31	2.6e+08	A/A3.H8	A/A3.H4'	1.49	5.00
142: P151	8.00	3.11	7.4e+07	A/A4.H8	A/A4.H4'	1.49	7.00
143: P152	7.90	2.42	2e+08	A/A5.H8	A/A5.H4'	1.49	5.00
144: P153	7.47	3.23	3.9e+07	A/A6.H8	A/A6.H4'	1.49	7.00
145: P154	7.18	4.18	2.1e+08	A/C8.H6	A/C8.H4'	1.49	5.00
146: P155	7.58	4.11	1.7e+08	A/G9.H8	A/G9.H4'	1.49	5.00
147: P157	7.38	3.91	3.9e+08	B/C11.H6	B/C11.H4'	1.49	3.00
148: P160	7.77	4.22	1.8e+08	B/G13.H8	B/G13.H4'	1.49	5.00
149: P161	7.46	4.24	2.2e+08	B/T14.H6	B/T14.H4'	1.49	5.00
150: P162	7.41	2.57	1.2e+08	B/T15.H6	B/T15.H4'	1.49	7.00
151: P163	7.18	3.97	8.8e+07	B/T16.H6	B/T16.H4'	1.49	7.00
152: P165	7.67	4.05	2e+08	B/G20.H8	B/G20.H4'	1.49	5.00
153: P164	7.03	2.16	2.1e+08	B/C19.H6	B/C19.H4'	1.49	5.00
154: P166	7.25	3.56	2e+08	A/C1.H6	A/C1.H5'1	1.49	5.00
155: P167	7.25	3.48	2e+08	A/C1.H6	A/C1.H5'2	1.49	5.00
156: P168	7.72	3.74	1.8e+08	A/G2.H8	A/G2.H5'2	1.49	5.00
157: P169	7.72	3.82	2e+08	A/G2.H8	A/G2.H5'1	1.49	5.00
158: P170	7.94	4.06	2.6e+08	A/A3.H8	A/A3.H5'2	1.49	5.00
159: P171	7.94	4.12	2.6e+08	A/A3.H8	A/A3.H5'1	1.49	5.00
160: P172	8.00	3.46	8e+07	A/A4.H8	A/A4.H5'2	1.49	7.00
161: P173	8.00	3.73	1e+08	A/A4.H8	A/A4.H5'1	1.49	7.00
162: P174	7.91	3.11	5.3e+07	A/A5.H8	A/A5.H5'2	1.49	7.00
163: P175	7.91	3.72	1.7e+08	A/A5.H8	A/A5.H5'1	1.49	5.00
164: P176	7.47	3.72	9.2e+07	A/A6.H8	A/A6.H5'1	1.49	7.00
165: P179	7.18	3.55	1e+08	A/C8.H6	A/C8.H5'2	1.49	7.00
166: P180	7.58	3.95	2e+08	A/G9.H8	A/G9.H5'2	1.49	5.00
167: P181	7.58	4.04	1.5e+08	A/G9.H8	A/G9.H5'1	1.49	5.00
168: P183	7.19	3.96	1.3e+08	A/C8.H6	A/C8.H5'1	1.49	7.00
169: P185	7.38	3.46	1.6e+08	B/C11.H6	B/C11.H5'2	1.49	5.00
170: P186	7.39	3.52	2e+08	B/C11.H6	B/C11.H5'1	1.49	5.00
171: P187	7.34	3.76	1.6e+08	B/C12.H6	B/C12.H5'2	1.49	5.00
172: P188	7.34	3.84	1.8e+08	B/C12.H6	B/C12.H5'1	1.49	5.00
173: P189	7.77	3.94	1.5e+08	B/G13.H8	B/G13.H5'1	1.49	5.00
174: P191	7.45	4.01	2e+08	B/T14.H6	B/T14.H5'2	1.49	5.00
175: P192	7.45	4.09	2.3e+08	B/T14.H6	B/T14.H5'1	1.49	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
176: P193	7.41	4.02	2.6e+07	B/T15.H6	B/T15.H5'2	1.49	7.00
177: P194	7.41	4.25	2.4e+08	B/T15.H6	B/T15.H5'1	1.49	5.00
178: P195	7.18	2.91	6.4e+07	B/T16.H6	B/T16.H5'2	1.49	7.00
179: P196	7.18	3.44	2.6e+07	B/T16.H6	B/T16.H5'1	1.49	7.00
180: P197	7.02	3.73	2.5e+08	B/T17.H6	B/T17.H5'1	1.49	5.00
181: P198	7.03	2.89	1.4e+08	B/C19.H6	B/C19.H5'2	1.49	7.00
182: P199	7.03	3.31	5.1e+07	B/C19.H6	B/C19.H5'1	1.49	7.00
183: P200	7.66	3.88	1.2e+08	B/G20.H8	B/G20.H5'2	1.49	7.00
184: P201	7.66	3.93	1.1e+08	B/G20.H8	B/G20.H5'1	1.49	7.00
185: P202	7.71	7.25	1.5e+08	A/G2.H8	A/C1.H6	1.49	5.00
186: P203	7.95	7.71	1.7e+08	A/A3.H8	A/G2.H8	1.49	5.00
187: P204	8.00	7.92	2.7e+08	A/A4.H8	A/A5.H8	1.49	5.00
188: P205	7.91	7.47	1.1e+08	A/A5.H8	A/A6.H8	1.49	7.00
189: P206	8.27	7.47	3.8e+06	A/A7.H8	A/A6.H8	1.49	7.00
190: P207	7.58	7.18	4.9e+06	A/G9.H8	A/C8.H6	1.49	7.00
191: P208	7.41	7.18	7.9e+07	B/T15.H6	B/T16.H6	1.49	7.00
192: P209	7.58	7.25	4.3e+07	A/G9.H8	A/G10.H8	1.49	7.00
193: P210	7.76	7.35	1.3e+08	B/G13.H8	B/C12.H6	1.49	7.00
194: P211	7.76	7.45	1.9e+08	B/G13.H8	B/T14.H6	1.49	5.00
195: P212	7.18	7.02	5.3e+07	B/T16.H6	B/T17.H6	1.49	7.00
196: P213	7.02	6.95	1.2e+08	B/T17.H6	B/T18.H6	1.49	7.00
197: P214	7.67	7.03	1e+08	B/G20.H8	B/C19.H6	1.49	7.00
198: P215	5.32	5.15	8.9e+07	A/C1.H1'	A/G2.H1'	1.49	7.00
199: P216	5.75	5.15	5.5e+07	A/A3.H1'	A/G2.H1'	1.49	7.00
200: P217	5.67	5.07	3.3e+07	A/A5.H1'	A/A6.H1'	1.49	7.00
201: P218	5.20	5.08	1e+08	A/C8.H1'	A/A7.H1'	1.49	7.00
202: P219	5.94	5.19	3.2e+07	A/G9.H1'	A/C8.H1'	1.49	7.00
203: P223	6.15	5.72	1.7e+07	B/T14.H1'	B/G13.H1'	1.49	7.00
204: P224	5.71	5.05	6.6e+07	B/T16.H1'	B/T17.H1'	1.49	7.00
205: P225	5.44	5.07	8.8e+07	B/T18.H1'	B/T17.H1'	1.49	7.00
206: P226	5.90	5.16	3.8e+07	B/G20.H1'	B/C19.H1'	1.49	7.00
207: P227	5.44	5.16	1.9e+08	B/T18.H1'	B/C19.H1'	1.49	5.00
208: P228	5.32	4.39	3.4e+08	A/C1.H1'	A/C1.H3'	1.49	3.00
209: P229	5.14	4.81	2.5e+08	A/G2.H1'	A/G2.H3'	1.49	5.00
210: P230	5.76	4.91	2.8e+08	A/A3.H1'	A/A3.H3'	1.49	3.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
211: P231	5.56	4.50	1.7e+08	A/A4.H1'	A/A4.H3'	1.49	5.00
212: P232	5.68	4.30	2.5e+08	A/A5.H1'	A/A5.H3'	1.49	5.00
213: P233	5.07	4.40	1.6e+08	A/A6.H1'	A/A6.H3'	1.49	5.00
214: P234	5.08	4.30	1.5e+08	A/A7.H1'	A/A7.H3'	1.49	5.00
215: P235	5.20	4.80	2.9e+08	A/C8.H1'	A/C8.H3'	1.49	3.00
216: P236	5.94	4.51	3.2e+08	A/G9.H1'	A/G9.H3'	1.49	3.00
217: P238	5.61	4.40	2.1e+08	B/C11.H1'	B/C11.H3'	1.49	5.00
218: P240	5.72	4.84	4.6e+08	B/G13.H1'	B/G13.H3'	1.49	3.00
219: P241	6.15	4.82	1.2e+08	B/T14.H1'	B/T14.H3'	1.49	7.00
220: P242	6.08	4.30	8.2e+07	B/T15.H1'	B/T15.H3'	1.49	7.00
221: P243	5.71	4.39	1e+08	B/T16.H1'	B/T16.H3'	1.49	7.00
222: P244	5.06	4.30	1.6e+08	B/T17.H1'	B/T17.H3'	1.49	5.00
223: P245	5.44	4.40	4.3e+08	B/T18.H1'	B/T18.H3'	1.49	3.00
224: P246	5.17	4.53	1.2e+08	B/C19.H1'	B/C19.H3'	1.49	7.00
225: P247	5.90	4.52	4.1e+08	B/G20.H1'	B/G20.H3'	1.49	3.00
226: P248	5.32	3.83	3.3e+08	A/C1.H1'	A/C1.H4'	1.49	3.00
227: P249	5.14	4.13	2.4e+08	A/G2.H1'	A/G2.H4'	1.49	5.00
228: P250	5.75	4.31	5.8e+08	A/A3.H1'	A/A3.H4'	1.49	3.00
229: P251	5.56	3.11	2.7e+07	A/A4.H1'	A/A4.H4'	1.49	7.00
230: P252	5.68	2.42	4.4e+08	A/A5.H1'	A/A5.H4'	1.49	3.00
231: P253	5.07	3.23	3.5e+06	A/A6.H1'	A/A6.H4'	1.49	7.00
232: P254	5.08	2.45	1.8e+08	A/A7.H1'	A/A7.H4'	1.49	5.00
233: P255	5.20	4.18	2.4e+08	A/C8.H1'	A/C8.H4'	1.49	5.00
234: P256	5.93	4.11	3.6e+08	A/G9.H1'	A/G9.H4'	1.49	3.00
235: P258	5.61	3.90	3.2e+08	B/C11.H1'	B/C11.H4'	1.49	3.00
236: P260	5.73	4.22	2.3e+08	B/G13.H1'	B/G13.H4'	1.49	5.00
237: P261	6.15	4.24	2.6e+08	B/T14.H1'	B/T14.H4'	1.49	5.00
238: P262	6.07	2.57	2.5e+08	B/T15.H1'	B/T15.H4'	1.49	5.00
239: P263	5.71	3.97	1.8e+08	B/T16.H1'	B/T16.H4'	1.49	5.00
240: P264	5.44	3.91	3.6e+08	B/T18.H1'	B/T18.H4'	1.49	3.00
241: P265	5.16	2.15	3.2e+08	B/C19.H1'	B/C19.H4'	1.49	3.00
242: P266	5.90	4.05	4.1e+08	B/G20.H1'	B/G20.H4'	1.49	3.00
243: P267	5.32	3.48	1.3e+08	A/C1.H1'	A/C1.H5'2	1.49	7.00
244: P268	5.33	3.56	1.5e+08	A/C1.H1'	A/C1.H5'1	1.49	5.00
245: P269	5.14	3.74	1.4e+08	A/G2.H1'	A/G2.H5'2	1.49	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
246: P270	5.14	3.81	1.4e+08	A/G2.H1'	A/G2.H5'1	1.49	5.00
247: P271	5.75	4.06	2.1e+08	A/A3.H1'	A/A3.H5'2	1.49	5.00
248: P272	5.75	4.12	2.1e+08	A/A3.H1'	A/A3.H5'1	1.49	5.00
249: P273	5.56	3.46	7.4e+07	A/A4.H1'	A/A4.H5'2	1.49	7.00
250: P274	5.56	3.73	9.5e+07	A/A4.H1'	A/A4.H5'1	1.49	7.00
251: P275	5.68	3.11	1e+08	A/A5.H1'	A/A5.H5'2	1.49	7.00
252: P276	5.68	3.72	2.2e+08	A/A5.H1'	A/A5.H5'1	1.49	5.00
253: P277	5.07	3.57	1.3e+08	A/A6.H1'	A/A6.H5'2	1.49	7.00
254: P278	5.07	3.72	1.4e+07	A/A6.H1'	A/A6.H5'1	1.49	7.00
255: P279	5.08	4.03	1.2e+08	A/A7.H1'	A/A7.H5'2	1.49	7.00
256: P280	5.08	4.19	1e+08	A/A7.H1'	A/A7.H5'1	1.49	7.00
257: P281	5.20	3.55	1.3e+08	A/C8.H1'	A/C8.H5'2	1.49	7.00
258: P282	5.20	3.96	3.4e+08	A/C8.H1'	A/C8.H5'1	1.49	3.00
259: P283	5.93	3.96	4.8e+08	A/G9.H1'	A/G9.H5'2	1.49	3.00
260: P284	5.94	4.04	4.2e+08	A/G9.H1'	A/G9.H5'1	1.49	3.00
261: P287	5.61	3.47	1.4e+08	B/C11.H1'	B/C11.H5'2	1.49	7.00
262: P288	5.61	3.52	1.9e+08	B/C11.H1'	B/C11.H5'1	1.49	5.00
263: P292	5.72	3.94	2.5e+08	B/G13.H1'	B/G13.H5'1	1.49	5.00
264: P293	6.15	4.01	2.3e+08	B/T14.H1'	B/T14.H5'2	1.49	5.00
265: P294	6.15	4.09	1.6e+08	B/T14.H1'	B/T14.H5'1	1.49	5.00
266: P295	6.08	4.02	2.2e+08	B/T15.H1'	B/T15.H5'2	1.49	5.00
267: P296	6.08	4.25	5.9e+07	B/T15.H1'	B/T15.H5'1	1.49	7.00
268: P297	5.71	3.45	1e+08	B/T16.H1'	B/T16.H5'1	1.49	7.00
269: P298	5.71	2.91	1.4e+08	B/T16.H1'	B/T16.H5'2	1.49	7.00
270: P299	5.06	3.44	5.8e+07	B/T17.H1'	B/T17.H5'2	1.49	7.00
271: P300	5.06	3.73	1.7e+08	B/T17.H1'	B/T17.H5'1	1.49	5.00
272: P301	5.44	3.61	6.7e+07	B/T18.H1'	B/T18.H5'1	1.49	7.00
273: P302	5.17	3.31	7.3e+07	B/C19.H1'	B/C19.H5'1	1.49	7.00
274: P303	5.17	2.89	1.4e+08	B/C19.H1'	B/C19.H5'2	1.49	7.00
275: P304	5.90	3.93	1.8e+08	B/G20.H1'	B/G20.H5'1	1.49	5.00
276: P305	5.90	3.90	2.4e+08	B/G20.H1'	B/G20.H5'2	1.49	5.00
277: P306	5.32	1.31	4e+08	A/C1.H1'	A/C1.H2'1	1.49	3.00
278: P307	5.32	2.01	5.9e+08	A/C1.H1'	A/C1.H2'2	1.49	3.00
279: P308	5.14	2.50	3.8e+08	A/G2.H1'	A/G2.H2'1	1.49	3.00
280: P309	5.14	2.55	3.5e+08	A/G2.H1'	A/G2.H2'2	1.49	3.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
316: P350	4.80	3.96	4.4e+08	A/C8.H3'	A/C8.H5'1	1.49	3.00
317: P351	4.80	4.19	3.4e+08	A/C8.H3'	A/C8.H4'	1.49	3.00
318: P352	4.66	3.91	3.4e+08	B/C12.H3'	B/C12.H4'	1.49	3.00
319: P353	4.66	3.77	1.8e+08	B/C12.H3'	B/C12.H5'2	1.49	5.00
320: P354	4.66	3.84	1.5e+07	B/C12.H3'	B/C12.H5'1	1.49	7.00
321: P355	4.30	3.12	6.1e+07	B/T17.H3'	B/T17.H4'	1.49	7.00
322: P356	4.30	3.44	3.2e+08	B/T17.H3'	B/T17.H5'2	1.49	3.00
323: P357	4.30	3.73	1.5e+09	B/T17.H3'	B/T17.H5'1	1.49	3.00
324: P358	4.30	3.10	6e+07	A/A5.H3'	A/A5.H5'2	1.49	7.00
325: P359	4.30	3.72	1.5e+08	A/A5.H3'	A/A5.H5'1	1.49	5.00
326: P360	4.30	2.42	3.3e+08	A/A5.H3'	A/A5.H4'	1.49	3.00
327: P361	4.49	3.11	2.3e+08	A/A4.H3'	A/A4.H4'	1.49	5.00
328: P362	4.49	3.46	4.4e+08	A/A4.H3'	A/A4.H5'2	1.49	3.00
329: P363	4.40	3.23	1.6e+08	A/A6.H3'	A/A6.H4'	1.49	5.00
330: P364	4.40	3.72	2.5e+08	A/A6.H3'	A/A6.H5'1	1.49	5.00
331: P365	4.40	3.58	1.9e+08	A/A6.H3'	A/A6.H5'2	1.49	5.00
332: P366	4.52	2.89	2.1e+08	B/C19.H3'	B/C19.H5'2	1.49	5.00
333: P367	4.53	2.16	2e+08	B/C19.H3'	B/C19.H4'	1.49	5.00
334: P368	4.41	3.23	2.6e+08	B/T18.H3'	B/T18.H5'2	1.49	5.00
335: P369	4.41	3.61	2e+08	B/T18.H3'	B/T18.H5'1	1.49	5.00
336: P370	4.41	3.91	3.5e+08	B/T18.H3'	B/T18.H4'	1.49	3.00
337: P371	4.39	2.92	2.9e+08	B/T16.H3'	B/T16.H5'2	1.49	3.00
338: P372	4.39	3.44	5e+08	B/T16.H3'	B/T16.H5'1	1.49	3.00
339: P373	4.39	3.48	4.2e+08	A/C1.H3'	A/C1.H5'2	1.49	3.00
340: P374	4.39	3.56	3.9e+08	A/C1.H3'	A/C1.H5'1	1.49	3.00
341: P375	4.39	3.83	9.1e+08	A/C1.H3'	A/C1.H4'	1.49	3.00
342: P376	4.51	4.11	3.1e+08	A/G9.H3'	A/G9.H4'	1.49	3.00
343: P377	4.51	4.04	3.3e+08	A/G9.H3'	A/G9.H5'1	1.49	3.00
344: P378	4.51	3.96	3.4e+08	A/G9.H3'	A/G9.H5'2	1.49	3.00
345: P379	4.81	3.74	2.4e+08	A/G2.H3'	A/G2.H5'2	1.49	5.00
346: P380	4.81	3.81	5.7e+08	A/G2.H3'	A/G2.H5'1	1.49	3.00
347: P381	4.81	4.13	4e+08	A/G2.H3'	A/G2.H4'	1.49	3.00
348: P383	4.83	4.24	3.3e+08	B/T14.H3'	B/T14.H4'	1.49	3.00
349: P386	4.85	3.94	2.9e+08	B/G13.H3'	B/G13.H5'1	1.49	3.00
350: P387	4.84	4.22	2.6e+08	B/G13.H3'	B/G13.H4'	1.49	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
281: P310	5.76	2.73	3.7e+08	A/A3.H1'	A/A3.H2'2	1.49	3.00
282: P311	5.76	2.52	3.3e+08	A/A3.H1'	A/A3.H2'1	1.49	3.00
283: P312	5.56	2.43	1.7e+08	A/A4.H1'	A/A4.H2'1	1.49	5.00
284: P313	5.56	2.45	1.7e+08	A/A4.H1'	A/A4.H2'2	1.49	5.00
285: P314	5.68	2.54	3.8e+08	A/A5.H1'	A/A5.H2'1	1.49	3.00
286: P315	5.68	2.90	1.9e+08	A/A5.H1'	A/A5.H2'2	1.49	5.00
287: P316	5.07	1.47	3.8e+08	A/A6.H1'	A/A6.H2'1	1.49	3.00
288: P317	5.07	1.82	2.7e+08	A/A6.H1'	A/A6.H2'2	1.49	3.00
289: P318	5.08	2.56	1.7e+08	A/A7.H1'	A/A7.H2'1	1.49	5.00
290: P319	5.08	2.57	2e+08	A/A7.H1'	A/A7.H2'2	1.49	5.00
291: P320	5.20	2.21	3.2e+08	A/C8.H1'	A/C8.H2'1	1.49	3.00
292: P321	5.20	2.46	4.2e+08	A/C8.H1'	A/C8.H2'2	1.49	3.00
293: P322	5.94	2.21	4.2e+08	A/G9.H1'	A/G9.H2'2	1.49	3.00
294: P323	5.94	2.39	4.1e+08	A/G9.H1'	A/G9.H2'1	1.49	3.00
295: P326	5.61	1.63	4.1e+08	B/C11.H1'	B/C11.H2'1	1.49	3.00
296: P327	5.61	2.21	6.7e+08	B/C11.H1'	B/C11.H2'2	1.49	3.00
297: P330	5.73	2.41	3e+08	B/G13.H1'	B/G13.H2'1	1.49	3.00
298: P331	5.73	2.55	3.2e+08	B/G13.H1'	B/G13.H2'2	1.49	3.00
299: P332	6.14	2.21	1.3e+08	B/T14.H1'	B/T14.H2'1	1.49	7.00
300: P333	6.15	2.55	2.5e+08	B/T14.H1'	B/T14.H2'2	1.49	5.00
301: P334	6.07	1.39	8.4e+07	B/T15.H1'	B/T15.H2'1	1.49	7.00
302: P335	6.08	1.60	1.7e+08	B/T15.H1'	B/T15.H2'2	1.49	5.00
303: P336	5.71	1.46	3.2e+08	B/T16.H1'	B/T16.H2'2	1.49	3.00
304: P337	5.71	2.19	3.8e+08	B/T16.H1'	B/T16.H2'1	1.49	3.00
305: P338	5.07	1.72	1e+08	B/T17.H1'	B/T17.H2'1	1.49	7.00
306: P339	5.06	1.84	2.7e+08	B/T17.H1'	B/T17.H2'2	1.49	5.00
307: P341	5.44	1.80	1.9e+08	B/T18.H1'	B/T18.H2'1	1.49	5.00
308: P342	5.17	1.49	2e+08	B/C19.H1'	B/C19.H2'2	1.49	5.00
309: P343	5.17	1.92	3.9e+08	B/C19.H1'	B/C19.H2'1	1.49	3.00
310: P344	5.90	2.19	4.4e+08	B/G20.H1'	B/G20.H2'2	1.49	3.00
311: P345	5.89	2.43	3.9e+08	B/G20.H1'	B/G20.H2'1	1.49	3.00
312: P346	4.92	4.31	2.9e+08	A/A3.H3'	A/A3.H4'	1.49	3.00
313: P347	4.91	4.12	3.2e+08	A/A3.H3'	A/A3.H5'1	1.49	3.00
314: P348	4.92	4.06	3.4e+08	A/A3.H3'	A/A3.H5'2	1.49	3.00
315: P349	4.80	3.55	2.3e+08	A/C8.H3'	A/C8.H5'2	1.49	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
351: P388	4.52	4.05	2.5e+08	B/G20.H3'	B/G20.H4'	1.49	5.00
352: P389	4.52	3.93	1.1e+08	B/G20.H3'	B/G20.H5'1	1.49	7.00
353: P390	4.52	3.89	9.2e+07	B/G20.H3'	B/G20.H5'2	1.49	7.00
354: P391	4.39	1.31	3.4e+08	A/C1.H3'	A/C1.H2'1	1.49	3.00
355: P392	4.39	2.01	2.5e+08	A/C1.H3'	A/C1.H2'2	1.49	5.00
356: P395	4.83	2.21	2e+08	B/T14.H3'	B/T14.H2'1	1.49	5.00
357: P396	4.83	2.55	4.3e+08	B/T14.H3'	B/T14.H2'2	1.49	3.00
358: P397	4.92	2.72	3.4e+08	A/A3.H3'	A/A3.H2'2	1.49	3.00
359: P398	4.92	2.52	4.5e+08	A/A3.H3'	A/A3.H2'1	1.49	3.00
360: P399	4.66	1.99	1.4e+08	B/C12.H3'	B/C12.H2'2	1.49	7.00
361: P400	4.66	2.20	2.6e+08	B/C12.H3'	B/C12.H2'1	1.49	5.00
362: P401	4.53	1.49	2.4e+08	B/C19.H3'	B/C19.H2'2	1.49	5.00
363: P402	4.53	1.92	2.2e+08	B/C19.H3'	B/C19.H2'1	1.49	5.00
364: P403	4.40	1.47	1.2e+08	A/A6.H3'	A/A6.H2'1	1.49	7.00
365: P404	4.40	1.82	1.6e+08	A/A6.H3'	A/A6.H2'2	1.49	5.00
366: P405	4.39	1.46	1.4e+08	B/T16.H3'	B/T16.H2'2	1.49	7.00
367: P406	4.39	2.19	1.4e+08	B/T16.H3'	B/T16.H2'1	1.49	7.00
368: P409	4.81	2.50	3.8e+08	A/G2.H3'	A/G2.H2'1	1.49	3.00
369: P410	4.81	2.56	1.7e+08	A/G2.H3'	A/G2.H2'2	1.49	5.00
370: P411	4.85	2.55	7e+08	B/G13.H3'	B/G13.H2'2	1.49	3.00
371: P412	4.85	2.41	4e+08	B/G13.H3'	B/G13.H2'1	1.49	3.00
372: P413	4.30	1.61	6e+07	B/T15.H3'	B/T15.H2'2	1.49	7.00
373: P414	4.30	1.38	6.5e+08	B/T15.H3'	B/T15.H2'1	1.49	3.00
374: P415	4.40	1.62	8.8e+07	B/C11.H3'	B/C11.H2'1	1.49	7.00
375: P416	4.40	2.21	1.4e+08	B/C11.H3'	B/C11.H2'2	1.49	5.00
376: P417	4.30	2.90	9.4e+07	A/A5.H3'	A/A5.H2'2	1.49	7.00
377: P418	4.30	2.54	1.9e+08	A/A5.H3'	A/A5.H2'1	1.49	5.00
378: P419	4.50	2.45	1e+08	A/A4.H3'	A/A4.H2'2	1.49	7.00
379: P420	4.50	2.43	1.1e+08	A/A4.H3'	A/A4.H2'1	1.49	7.00
380: P421	4.30	2.56	1.2e+07	A/A7.H3'	A/A7.H2'1	1.49	7.00
381: P422	4.30	2.57	1.2e+08	A/A7.H3'	A/A7.H2'2	1.49	7.00
382: P423	4.52	2.20	1.9e+08	B/G20.H3'	B/G20.H2'2	1.49	5.00
383: P424	4.52	2.43	2.3e+08	B/G20.H3'	B/G20.H2'1	1.49	5.00
384: P425	4.51	2.21	2.4e+08	A/G9.H3'	A/G9.H2'2	1.49	5.00
385: P426	4.51	2.39	2.7e+08	A/G9.H3'	A/G9.H2'1	1.49	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
386: P427	4.80	2.21	2.2e+08	A/C8.H3'	A/C8.H2'1	1.49	5.00
387: P428	4.80	2.46	4e+08	A/C8.H3'	A/C8.H2'2	1.49	3.00
388: P429	4.30	1.83	2.4e+08	B/T17.H3'	B/T17.H2'2	1.49	5.00
389: P430	4.30	1.73	7.9e+07	B/T17.H3'	B/T17.H2'1	1.49	7.00
390: P432	4.41	1.80	1.7e+08	B/T18.H3'	B/T18.H2'1	1.49	5.00
391: P433	3.57	1.32	1.7e+08	A/C1.H5'1	A/C1.H2'1	1.49	5.00
392: P434	3.49	1.31	1.6e+08	A/C1.H5'2	A/C1.H2'1	1.49	5.00
393: P435	3.53	1.63	1.1e+08	B/C11.H5'1	B/C11.H2'1	1.49	7.00
394: P436	3.46	1.63	1.2e+08	B/C11.H5'2	B/C11.H2'1	1.49	7.00
395: P437	3.57	2.01	2.3e+08	A/C1.H5'1	A/C1.H2'2	1.49	5.00
396: P438	3.49	2.01	2.4e+08	A/C1.H5'2	A/C1.H2'2	1.49	5.00
397: P439	3.52	2.21	3.5e+08	B/C11.H5'1	B/C11.H2'2	1.49	3.00
398: P440	3.46	2.21	2.6e+08	B/C11.H5'2	B/C11.H2'2	1.49	5.00
399: P441	3.46	2.44	2.4e+08	A/A4.H5'2	A/A4.H2'1	1.49	5.00
400: P442	3.46	2.45	2.2e+07	A/A4.H5'2	A/A4.H2'2	1.49	7.00
401: P443	3.44	1.84	2.9e+08	B/T17.H5'2	B/T17.H2'2	1.49	3.00
402: P444	3.73	1.84	5e+08	B/T17.H5'1	B/T17.H2'2	1.49	3.00
403: P445	3.73	1.72	3.1e+08	B/T17.H5'1	B/T17.H2'1	1.49	3.00
404: P446	3.73	1.47	4e+08	A/A6.H5'1	A/A6.H2'1	1.49	3.00
405: P447	3.72	1.82	2.7e+08	A/A6.H5'1	A/A6.H2'2	1.49	5.00
406: P448	2.91	1.46	1.7e+08	B/T16.H5'2	B/T16.H2'2	1.49	5.00
407: P449	3.11	2.42	3.7e+08	A/A5.H5'2	A/A5.H4'	1.49	3.00
408: P451	2.57	1.39	3.6e+08	B/T15.H4'	B/T15.H2'1	1.49	3.00
409: P452	3.83	1.30	3.5e+08	A/C1.H4'	A/C1.H2'1	1.49	3.00
410: P453	3.83	2.01	6.6e+08	A/C1.H4'	A/C1.H2'2	1.49	3.00
411: P457	3.97	1.46	1.8e+08	B/T16.H4'	B/T16.H2'2	1.49	5.00
412: P458	4.11	2.21	3.7e+08	A/G9.H4'	A/G9.H2'2	1.49	3.00
413: P459	3.31	2.89	6.1e+08	B/C19.H5'1	B/C19.H5'2	1.49	3.00
414: P460	3.46	3.11	2.1e+08	A/A4.H5'2	A/A4.H4'	1.49	5.00
415: P461	3.72	3.11	2.6e+08	A/A5.H5'1	A/A5.H5'2	1.49	5.00
416: P462	3.73	3.11	7.3e+07	A/A4.H5'1	A/A4.H4'	1.49	7.00
417: P463	3.72	2.90	2e+08	A/A5.H5'1	A/A5.H2'2	1.49	5.00
418: P464	4.19	3.55	3.6e+08	A/C8.H4'	A/C8.H5'2	1.49	3.00
419: P465	2.02	1.31	3.1e+08	A/C1.H2'2	A/C1.H2'1	1.49	3.00
420: P467	2.19	1.46	6.5e+08	B/T16.H2'1	B/T16.H2'2	1.49	3.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
421: P468	1.82	1.47	7.7e+08	A/A6.H2'2	A/A6.H2'1	1.49	3.00
422: P469	2.90	2.54	1.8e+08	A/A5.H2'2	A/A5.H2'1	1.49	5.00
423: P470	2.55	2.21	7.4e+08	B/T14.H2'2	B/T14.H2'1	1.49	3.00
424: P471	2.20	1.99	2.6e+08	B/C12.H2'1	B/C12.H2'2	1.49	5.00
425: P473	1.91	1.49	5.2e+08	B/C19.H2'1	B/C19.H2'2	1.49	3.00
426: P474	2.21	1.62	3.9e+08	B/C11.H2'2	B/C11.H2'1	1.49	3.00
427: P475	8.11	7.63	3.1e+08	A/OH1.H6'1	A/OH1.H7'1	1.49	3.00
428: P476	7.82	7.48	5.6e+08	A/OH1.H4'2	A/OH1.H5'2	1.49	3.00
429: P477	7.47	7.33	9.2e+08	A/OH1.H5'2	A/OH1.H6'2	1.49	3.00
430: P478	8.00	7.32	3.6e+08	A/OH1.H7'2	A/OH1.H6'2	1.49	3.00
431: P479	7.90	7.82	2e+08	A/OH1.H3'2	A/OH1.H4'2	1.49	5.00
432: P480	8.00	7.82	1.7e+08	A/OH1.H7'2	A/OH1.H4'2	1.49	5.00
433: P482	8.00	7.47	1.6e+08	A/OH1.H7'2	A/OH1.H5'2	1.49	5.00
434: P483	8.11	7.90	1.4e+07	A/OH1.H6'1	A/OH1.H3'2	1.49	7.00
435: P484	7.76	7.44	3.6e+08	A/OH1.H4'1	A/OH1.H3'1	1.49	3.00
436: P485	7.76	8.11	3.2e+07	A/OH1.H4'1	A/OH1.H6'1	1.49	7.00
437: P487	7.90	7.47	1.1e+08	A/OH1.H3'2	A/OH1.H5'2	1.49	7.00
438: P488	7.83	7.31	4e+08	A/OH1.H4'2	A/OH1.H6'2	1.49	3.00
439: P489	7.90	7.03	3.4e+07	A/OH1.H3	A/OH1.H6	1.49	7.00
440: P490	7.90	7.44	6.6e+07	A/OH1.H3	A/OH1.H3'1	1.49	7.00
441: P491	7.90	5.15	1.4e+08	A/OH1.H3	A/OH1.H8B	1.49	7.00
442: P492	7.90	3.73	6.3e+07	A/OH1.H3	A/OH1.H8A	1.49	7.00
443: P493	7.45	4.81	6.4e+07	A/OH1.H3'1	A/OH1.H1B	1.49	7.00
444: P494	5.15	4.81	1.9e+08	A/OH1.H8B	A/OH1.H1B	1.49	5.00
445: P495	5.15	4.30	3.2e+08	A/OH1.H8B	A/OH1.H8A2	1.49	3.00
446: P496	5.15	3.73	1.8e+08	A/OH1.H8B	A/OH1.H8A	1.49	5.00
447: P497	5.15	3.71	9.2e+07	A/OH1.H8B	A/OH1.H1A	1.49	7.00
448: P498	4.81	4.30	4.1e+08	A/OH1.H1B	A/OH1.H8A2	1.49	3.00
449: P499	4.81	3.71	5.2e+08	A/OH1.H1B	A/OH1.H1A	1.49	3.00
450: P500	4.81	3.73	2.4e+08	A/OH1.H1B	A/OH1.H8A	1.49	5.00
451: P501	4.30	3.71	1.3e+08	A/OH1.H8A2	A/OH1.H1A	1.49	7.00
452: P502	4.30	3.73	9.1e+07	A/OH1.H8A2	A/OH1.H8A	1.49	7.00
453: P503	8.16	7.68	2.1e+07	A/A7.H2	A/A6.H2	1.49	7.00
454: P504	8.11	7.68	3.1e+07	A/A5.H2	A/A6.H2	1.49	7.00
455: P505	7.55	7.31	2.2e+08	A/A3.H2	A/A4.H2	1.49	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
456: P506	7.31	7.02	3.1e+07	A/A4.H2	B/T17.H6	1.49	7.00
457: P507	7.53	7.03	2.5e+07	A/A3.H2	B/C19.H6	1.49	7.00
458: P508	7.68	7.47	4.3e+07	A/A6.H2	A/A6.H8	1.49	7.00
459: P509	8.16	5.44	3.5e+07	A/A7.H2	A/C8.H5	1.49	7.00
460: P510	8.12	5.06	5.1e+06	A/A5.H2	B/T17.H1'	1.49	7.00
461: P511	8.11	5.68	2.8e+07	A/A5.H2	A/A5.H1'	1.49	7.00
462: P512	7.68	5.71	1.1e+08	A/A6.H2	B/T16.H1'	1.49	7.00
463: P514	7.68	5.07	1.1e+08	A/A6.H2	A/A6.H1'	1.49	7.00
464: P515	7.55	5.90	1.6e+08	A/A3.H2	B/G20.H1'	1.49	5.00
465: P516	7.54	5.44	8.7e+07	A/A3.H2	B/T18.H1'	1.49	7.00
466: P517	7.31	5.44	3e+07	A/A4.H2	B/T18.H1'	1.49	7.00
467: P518	7.31	5.06	5.2e+07	A/A4.H2	B/T17.H1'	1.49	7.00
468: P524	7.68	3.12	1.3e+08	A/A6.H2	B/T17.H4'	1.49	7.00
469: P526	7.68	2.91	1.1e+08	A/A6.H2	B/T16.H5'2	1.49	7.00
470: P527	7.68	3.43	1.7e+08	A/A6.H2	B/T17.H5'2	1.49	5.00
471: P529	7.55	2.89	3.4e+07	A/A3.H2	B/C19.H5'2	1.49	7.00
472: P530	7.31	3.46	1.3e+08	A/A4.H2	A/A4.H5'2	1.49	7.00
473: P531	7.31	3.31	6.2e+07	A/A4.H2	B/C19.H5'1	1.49	7.00
474: P532	7.31	2.89	1.2e+08	A/A4.H2	B/C19.H5'2	1.49	7.00
475: P534	8.11	2.42	1.1e+07	A/A5.H2	A/A5.H4'	1.49	7.00
476: P535	7.31	2.45	1e+08	A/A4.H2	A/A4.H2'2	1.49	7.00
477: P536	7.31	2.43	8.5e+07	A/A4.H2	A/A4.H2'1	1.49	7.00
478: P537	7.68	1.81	1.4e+08	A/A6.H2	A/A6.H2'2	1.49	7.00
479: P538	7.68	1.47	6.9e+07	A/A6.H2	A/A6.H2'1	1.49	7.00
480: P539	7.68	1.38	9.6e+07	A/A6.H2	B/T15.H2'1	1.49	7.00
481: P540	8.11	1.72	9.7e+06	A/A5.H2	B/T17.H2'1	1.49	7.00
482: P541	8.12	1.83	2e+07	A/A5.H2	B/T17.H2'2	1.49	7.00
483: P542	7.55	1.92	9.4e+07	A/A3.H2	B/C19.H2'1	1.49	7.00
484: P543	7.03	5.36	4e+08	B/C19.H6	B/C19.H5	1.49	3.00
485: P544	6.95	5.36	6.7e+07	B/T18.H6	B/C19.H5	1.49	7.00
486: P545	7.39	5.52	1e+09	B/C11.H6	B/C11.H5	1.49	3.00
487: P546	7.26	5.43	8.9e+08	A/C1.H6	A/C1.H5	1.49	3.00
488: P547	7.18	5.44	6.6e+07	A/C8.H6	A/C8.H5	1.49	7.00
489: P548	7.34	5.44	6.8e+08	B/C12.H6	B/C12.H5	1.49	3.00
490: P549	7.39	5.44	2.2e+08	B/C11.H6	B/C12.H5	1.49	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
491: P550	5.37	5.16	1.4e+08	B/C19.H5	B/C19.H1'	1.49	5.00
492: P552	5.61	5.52	2.2e+08	B/C11.H1'	B/C11.H5	1.49	5.00
493: P553	5.44	5.32	1.4e+08	A/C1.H5	A/C1.H1'	1.49	7.00
494: P554	5.61	5.44	1.3e+08	B/C11.H1'	B/C12.H5	1.49	7.00
495: P555	5.44	5.08	1.6e+08	A/C8.H5	A/A7.H1'	1.49	5.00
496: P557	5.44	5.21	4.3e+07	A/C8.H5	A/C8.H1'	1.49	7.00
497: P558	5.44	3.47	1.6e+08	A/C1.H5	A/C1.H5'2	1.49	5.00
498: P559	5.44	3.56	1.1e+08	A/C1.H5	A/C1.H5'1	1.49	7.00
499: P560	5.52	3.46	1.2e+08	B/C11.H5	B/C11.H5'2	1.49	7.00
500: P561	5.52	3.53	1.6e+08	B/C11.H5	B/C11.H5'1	1.49	5.00
501: P562	5.52	3.91	1.7e+08	B/C11.H5	B/C11.H4'	1.49	5.00
502: P563	5.52	4.40	1.8e+08	B/C11.H5	B/C11.H3'	1.49	5.00
503: P564	5.44	4.39	2.3e+08	A/C1.H5	A/C1.H3'	1.49	5.00
504: P565	5.44	3.83	2.5e+08	A/C1.H5	A/C1.H4'	1.49	5.00
505: P567	5.44	3.55	1.2e+08	A/C8.H5	A/C8.H5'2	1.49	7.00
506: P568	5.44	3.96	6e+07	A/C8.H5	A/C8.H5'1	1.49	7.00
507: P569	5.44	4.19	8.1e+07	A/C8.H5	A/C8.H4'	1.49	7.00
508: P570	5.44	4.80	2.1e+07	A/C8.H5	A/C8.H3'	1.49	7.00
509: P571	5.44	3.76	8.7e+07	B/C12.H5	B/C12.H5'2	1.49	7.00
510: P572	5.44	3.91	2.3e+08	B/C12.H5	B/C12.H4'	1.49	5.00
511: P573	5.44	4.65	2.3e+07	B/C12.H5	B/C12.H3'	1.49	7.00
512: P574	5.37	4.52	1.2e+08	B/C19.H5	B/C19.H3'	1.49	7.00
513: P575	5.37	2.89	7.8e+07	B/C19.H5	B/C19.H5'2	1.49	7.00
514: P576	5.37	3.30	2.6e+07	B/C19.H5	B/C19.H5'1	1.49	7.00
515: P577	5.37	2.14	2.7e+08	B/C19.H5	B/C19.H4'	1.49	5.00
516: P578	5.52	1.62	3.1e+08	B/C11.H5	B/C11.H2'1	1.49	3.00
517: P579	5.52	2.21	3e+08	B/C11.H5	B/C11.H2'2	1.49	3.00
518: P580	5.44	2.01	1.5e+08	A/C1.H5	A/C1.H2'2	1.49	5.00
519: P581	5.44	1.31	1.3e+08	A/C1.H5	A/C1.H2'1	1.49	7.00
520: P582	5.38	1.49	2.1e+08	B/C19.H5	B/C19.H2'2	1.49	5.00
521: P583	5.37	1.93	1.9e+08	B/C19.H5	B/C19.H2'1	1.49	5.00
522: P584	5.44	2.21	2e+08	A/C8.H5	A/C8.H2'1	1.49	5.00
523: P585	5.44	1.99	2e+08	B/C12.H5	B/C12.H2'2	1.49	5.00
524: P586	5.44	2.20	2.1e+08	B/C12.H5	B/C12.H2'1	1.49	5.00
525: P587	5.44	2.46	2e+08	A/C8.H5	A/C8.H2'2	1.49	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
526: P588	7.77	1.62	1.7e+08	B/G13.H8	B/T14.C5M	1.49	5.00
527: P589	7.45	1.62	6.8e+08	B/T14.H6	B/T14.C5M	1.49	3.00
528: P590	7.41	1.47	3.1e+08	B/T15.H6	B/T15.C5M	1.49	3.00
529: P591	7.46	1.47	1.3e+08	B/T14.H6	B/T15.C5M	1.49	7.00
530: P592	7.18	1.60	2.4e+08	B/T16.H6	B/T16.C5M	1.49	5.00
531: P593	7.41	1.60	1.2e+08	B/T15.H6	B/T16.C5M	1.49	7.00
532: P594	7.18	1.37	1.3e+08	B/T16.H6	B/T17.C5M	1.49	7.00
533: P595	7.01	1.37	1.4e+08	B/T17.H6	B/T17.C5M	1.49	7.00
534: P596	6.95	1.35	2.3e+08	B/T18.H6	B/T18.C5M	1.49	5.00
535: P597	7.02	1.35	1.8e+08	B/T17.H6	B/T18.C5M	1.49	5.00
536: P598	6.07	1.47	9.1e+07	B/T15.H1'	B/T15.C5M	1.49	7.00
537: P599	6.15	1.47	1.3e+08	B/T14.H1'	B/T15.C5M	1.49	7.00
538: P600	6.15	1.62	3.1e+08	B/T14.H1'	B/T14.C5M	1.49	3.00
539: P601	5.73	1.62	1.3e+08	B/G13.H1'	B/T14.C5M	1.49	7.00
540: P602	5.71	1.58	4.1e+08	B/T16.H1'	B/T16.C5M	1.49	3.00
541: P603	6.07	1.58	1e+08	B/T15.H1'	B/T16.C5M	1.49	7.00
542: P604	5.71	1.38	1.2e+08	B/T16.H1'	B/T17.C5M	1.49	7.00
543: P605	5.06	1.38	1.1e+08	B/T17.H1'	B/T17.C5M	1.49	7.00
544: P606	5.44	1.35	1.2e+08	B/T18.H1'	B/T18.C5M	1.49	7.00
545: P607	5.06	1.35	2.5e+08	B/T17.H1'	B/T18.C5M	1.49	5.00
546: P608	4.84	1.62	9.6e+07	B/G13.H3'	B/T14.C5M	1.49	7.00
547: P609	4.24	1.62	1e+08	B/T14.H4'	B/T14.C5M	1.49	7.00
548: P610	4.39	1.58	2.4e+08	B/T16.H3'	B/T16.C5M	1.49	5.00
549: P611	4.41	1.34	9.7e+07	B/T18.H3'	B/T18.C5M	1.49	7.00
550: P613	4.30	1.38	3.6e+07	B/T17.H3'	B/T17.C5M	1.49	7.00
551: P614	8.11	7.01	1.3e+07	A/A5.H2	B/T17.H6	1.49	7.00
552: P615	8.11	7.00	1.2e+07	A/OH1.H6'1	B/T17.H6	1.49	7.00
553: P617	7.76	7.01	1.4e+07	A/OH1.H4'1	B/T17.H6	1.49	7.00
554: P618	7.47	7.02	1.5e+07	A/OH1.H5'2	B/T17.H6	1.49	7.00
555: P621	7.54	7.31	8.1e+07	A/A3.H2	A/OH1.H6'2	1.49	7.00
556: P622	7.67	7.45	4.1e+07	A/A6.H2	A/OH1.H3'1	1.49	7.00
557: P623	7.90	7.55	9e+07	A/OH1.H3'2	A/A3.H2	1.49	7.00
558: P624	7.83	7.55	1.8e+08	A/OH1.H4'2	A/A3.H2	1.49	5.00
559: P626	7.90	7.68	9.5e+07	A/OH1.H3	A/A6.H2	1.49	7.00
560: P627	7.68	7.75	1.7e+08	A/A6.H2	A/OH1.H4'1	1.49	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
561: P629	7.82	7.31	3.9e+08	A/OH1.H4'2	A/A4.H2	1.49	3.00
562: P630	7.45	6.07	4.4e+07	A/OH1.H3'1	B/T15.H1'	1.49	7.00
563: P631	8.16	5.09	3.5e+07	A/A7.H2	A/A7.H1'	1.49	7.00
564: P632	8.00	5.67	1.1e+08	A/OH1.H7'2	A/A5.H1'	1.49	7.00
565: P633	7.76	5.06	1.5e+08	A/OH1.H4'1	B/T17.H1'	1.49	5.00
566: P634	7.82	5.16	1.5e+08	A/OH1.H4'2	B/C19.H1'	1.49	5.00
567: P635	8.00	5.16	1.9e+08	A/OH1.H7'2	B/C19.H1'	1.49	5.00
568: P636	7.82	5.55	5.5e+07	A/OH1.H4'2	A/A4.H1'	1.49	7.00
569: P637	7.82	5.67	1.8e+08	A/OH1.H4'2	A/A5.H1'	1.49	5.00
570: P638	7.75	5.71	1.3e+08	A/OH1.H4'1	B/T16.H1'	1.49	7.00
571: P639	7.62	5.06	1e+08	A/OH1.H7'1	B/T17.H1'	1.49	7.00
572: P640	7.45	5.07	6.1e+07	A/OH1.H3'1	A/A6.H1'	1.49	7.00
573: P641	7.47	5.16	1.2e+08	A/OH1.H5'2	B/C19.H1'	1.49	7.00
574: P642	7.31	5.67	1.7e+08	A/OH1.H6'2	A/A5.H1'	1.49	5.00
575: P644	7.31	5.16	8.5e+07	A/OH1.H6'2	B/C19.H1'	1.49	7.00
576: P645	7.31	5.54	9.7e+07	A/OH1.H6'2	A/A4.H1'	1.49	7.00
577: P648	7.68	4.81	1.7e+08	A/A6.H2	A/OH1.H1B	1.49	5.00
578: P649	8.00	4.40	8.3e+07	A/OH1.H7'2	B/T18.H3'	1.49	7.00
579: P653	7.82	4.48	1.2e+08	A/OH1.H4'2	A/A4.H3'	1.49	7.00
580: P654	7.82	4.30	7.2e+07	A/OH1.H4'2	A/A5.H3'	1.49	7.00
581: P655	7.76	4.30	1.2e+08	A/OH1.H4'1	B/T17.H3'	1.49	7.00
582: P656	7.68	4.30	1e+08	A/A6.H2	A/OH1.H8A2	1.49	7.00
583: P657	7.62	4.30	1.1e+08	A/OH1.H7'1	B/T17.H3'	1.49	7.00
584: P659	7.47	4.48	1.2e+08	A/OH1.H5'2	A/A4.H3'	1.49	7.00
585: P660	7.03	4.51	8.6e+07	A/OH1.H6	A/G9.H3'	1.49	7.00
586: P661	8.16	3.71	6.3e+07	A/A7.H2	A/OH1.H1A	1.49	7.00
587: P662	8.16	3.73	7.2e+07	A/A7.H2	A/OH1.H8A	1.49	7.00
588: P658	8.11	3.73	1.4e+08	A/OH1.H6'1	B/T17.H5'1	1.49	7.00
589: P663	8.00	3.72	1.1e+08	A/OH1.H7'2	A/A5.H5'1	1.49	7.00
590: P664	7.82	3.72	1.2e+08	A/OH1.H4'2	A/A5.H5'1	1.49	7.00
591: P666	7.76	3.73	9.7e+07	A/OH1.H4'1	B/T17.H5'1	1.49	7.00
592: P667	7.68	3.71	9.8e+07	A/A6.H2	A/OH1.H1A	1.49	7.00
593: P668	7.68	3.73	9.9e+07	A/A6.H2	A/OH1.H8A	1.49	7.00
594: P669	7.47	3.72	9.5e+07	A/OH1.H5'2	A/A5.H5'1	1.49	7.00
595: P671	7.62	3.73	2.6e+08	A/OH1.H7'1	B/T17.H5'1	1.49	5.00

	1: POS1	2: POS2	3: INTEGRAL	4: ATOM_POS1	5: ATOM_POS2	6: LOWER	7: UPPER
596: P670	8.11	3.12	3.3e+07	A/OH1.H6'1	B/T17.H4'	1.49	7.00
597: P673	8.00	2.89	1.7e+08	A/OH1.H7'2	B/C19.H5'2	1.49	5.00
598: P674	8.00	3.30	3.9e+07	A/OH1.H7'2	B/C19.H5'1	1.49	7.00
599: P675	7.90	2.90	1.6e+08	A/OH1.H3'2	A/A5.H2'2	1.49	5.00
600: P676	7.90	3.23	4.9e+07	A/OH1.H3'2	B/T18.H5'2	1.49	7.00
601: P678	7.90	3.61	7.8e+07	A/OH1.H3'2	B/T18.H5'1	1.49	7.00
602: P679	7.82	2.90	8.1e+07	A/OH1.H4'2	A/A5.H2'2	1.49	7.00
603: P680	7.82	2.88	7e+07	A/OH1.H4'2	B/C19.H5'2	1.49	7.00
604: P681	7.82	3.11	5.3e+07	A/OH1.H4'2	A/A5.H5'2	1.49	7.00
605: P682	7.82	3.30	5.3e+07	A/OH1.H4'2	B/C19.H5'1	1.49	7.00
606: P683	7.82	3.46	1.2e+08	A/OH1.H4'2	A/A4.H5'2	1.49	7.00
607: P684	7.75	3.23	3.5e+07	A/OH1.H4'1	B/T18.H5'2	1.49	7.00
608: P685	7.75	3.44	1.2e+08	A/OH1.H4'1	B/T17.H5'2	1.49	7.00
609: P686	7.62	3.61	2e+08	A/OH1.H7'1	B/T18.H5'1	1.49	5.00
610: P687	7.62	3.12	2.1e+08	A/OH1.H7'1	B/T17.H4'	1.49	5.00
611: P688	7.63	3.44	1.6e+08	A/OH1.H7'1	B/T17.H5'2	1.49	5.00
612: P689	7.47	2.88	5.6e+07	A/OH1.H5'2	B/C19.H5'2	1.49	7.00
613: P690	7.47	2.90	4.3e+07	A/OH1.H5'2	A/A5.H2'2	1.49	7.00
614: P691	7.45	3.44	1.1e+08	A/OH1.H3'1	B/T17.H5'2	1.49	7.00
615: P692	7.47	3.31	4e+07	A/OH1.H5'2	B/C19.H5'1	1.49	7.00
616: P693	7.45	3.12	1.9e+07	A/OH1.H3'1	B/T17.H4'	1.49	7.00
617: P694	7.48	3.11	1.7e+07	A/OH1.H5'2	A/A5.H5'2	1.49	7.00
618: P695	7.31	2.89	1.9e+08	A/OH1.H6'2	B/C19.H5'2	1.49	5.00
619: P696	7.31	3.11	6.7e+07	A/OH1.H6'2	A/A5.H5'2	1.49	7.00
620: P697	7.31	3.30	4.8e+07	A/OH1.H6'2	B/C19.H5'1	1.49	7.00
621: P698	7.31	3.46	1.8e+08	A/OH1.H6'2	A/A4.H5'2	1.49	5.00
622: P699	8.00	2.42	4.6e+06	A/OH1.H7'2	A/A5.H4'	1.49	7.00
623: P700	7.82	2.45	1.2e+08	A/OH1.H4'2	A/A4.H2'2	1.49	7.00
624: P701	7.82	2.44	1.2e+08	A/OH1.H4'2	A/A4.H2'1	1.49	7.00
625: P704	7.62	1.84	1.3e+08	A/OH1.H7'1	B/T17.H2'2	1.49	7.00
626: P705	7.45	1.38	5.7e+07	A/OH1.H3'1	B/T15.H2'1	1.49	7.00
627: P709	8.00	1.92	6e+07	A/OH1.H7'2	B/C19.H2'1	1.49	7.00
628: P712	8.00	1.49	1.7e+07	A/OH1.H7'2	B/C19.H2'2	1.49	7.00
629: P713	5.72	2.90	1.2e+08	B/G13.H1'	A/OH1.C34	1.49	7.00
630: P714	5.19	2.90	2.1e+08	A/C8.H1'	A/OH1.C34	1.49	5.00

APPENDIX XVIV

The Master Spreadsheet for the adozelesin-5′-($C^1G^2A^3A^4A^5A^6A^{7*}C^8G^9G^{10}$)·5′-($C^{11}C^{12}G^{13}T^{14}T^{15}T^{16}T^{17}T^{18}C^{19}G^{20}$) DNA adduct.

Table 9 The master spreadsheet.

	1: SHIFT
1: A/C1.H5'1	3.561
2: A/C1.H5'2	3.481
3: A/C1.H4'	3.833
4: A/C1.H1'	5.319
5: A/C1.H2'1	1.308
6: A/C1.H2'2	2.010
7: A/C1.H3'	4.392
8: A/C1.H6	7.257
9: A/C1.H5	5.436
10: A/C1.H41	?
11: A/C1.H42	?
12: A/G2.H5'1	3.816
13: A/G2.H5'2	3.744
14: A/G2.H4'	4.128
15: A/G2.H1'	5.139
16: A/G2.H2'1	2.504
17: A/G2.H2'2	2.557
18: A/G2.H3'	4.812
19: A/G2.H8	7.715
20: A/G2.H1	?
21: A/G2.H21	?
22: A/G2.H22	?
23: A/A3.H5'1	4.119
24: A/A3.H5'2	4.055
25: A/A3.H4'	4.306
26: A/A3.H1'	5.744
27: A/A3.H2'1	2.520
28: A/A3.H2'2	2.727
29: A/A3.H3'	4.905
30: A/A3.H8	7.945
31: A/A3.H61	?
32: A/A3.H62	?
33: A/A3.H2	7.551
34: A/A4.H5'1	3.729
35: A/A4.H5'2	3.457

	1: SHIFT
36: A/A4.H4'	3.111
37: A/A4.H1'	5.551
38: A/A4.H2'1	2.435
39: A/A4.H2'2	2.451
40: A/A4.H3'	4.498
41: A/A4.H8	8.000
42: A/A4.H61	?
43: A/A4.H62	?
44: A/A4.H2	7.307
45: A/A5.H5'1	3.718
46: A/A5.H5'2	3.109
47: A/A5.H4'	2.424
48: A/A5.H1'	5.672
49: A/A5.H2'1	2.538
50: A/A5.H2'2	2.896
51: A/A5.H3'	4.299
52: A/A5.H8	7.904
53: A/A5.H61	?
54: A/A5.H62	?
55: A/A5.H2	8.111
56: A/A6.H5'1	3.723
57: A/A6.H5'2	3.577
58: A/A6.H4'	3.226
59: A/A6.H1'	5.072
60: A/A6.H2'1	1.473
61: A/A6.H2'2	1.816
62: A/A6.H3'	4.400
63: A/A6.H8	7.473
64: A/A6.H61	?
65: A/A6.H62	?
66: A/A6.H2	7.679
67: A/A7.H5'1	4.188
68: A/A7.H5'2	4.026
69: A/A7.H4'	2.450
70: A/A7.H1'	5.078

	1: SHIFT
71: A/A7.H2'1	2.562
72: A/A7.H2'2	2.565
73: A/A7.H3'	4.304
74: A/A7.H8	8.268
75: A/A7.H62	?
76: A/A7.H2	8.159
77: A/C8.H5'1	3.962
78: A/C8.H5'2	3.550
79: A/C8.H4'	4.187
80: A/C8.H1'	5.194
81: A/C8.H2'1	2.209
82: A/C8.H2'2	2.461
83: A/C8.H3'	4.798
84: A/C8.H6	7.183
85: A/C8.H5	5.443
86: A/C8.H41	?
87: A/C8.H42	?
88: A/G9.H5'1	4.045
89: A/G9.H5'2	3.957
90: A/G9.H4'	4.110
91: A/G9.H1'	5.925
92: A/G9.H2'1	2.387
93: A/G9.H2'2	2.212
94: A/G9.H3'	4.514
95: A/G9.H8	7.583
96: A/G9.H1	?
97: A/G9.H21	?
98: A/G9.H22	?
99: A/G10.H5'1	3.965
100: A/G10.H5'2	3.755
101: A/G10.H4'	4.108
102: A/G10.H1'	5.432
103: A/G10.H2'1	1.298
104: A/G10.H2'2	2.010
105: A/G10.H3'	4.817

	1: SHIFT
106: A/G10.H8	7.250
107: A/G10.H1	?
108: A/G10.H21	?
109: A/G10.H22	?
110: B/C11.H5'1	3.521
111: B/C11.H5'2	3.463
112: B/C11.H4'	3.908
113: B/C11.H1'	5.612
114: B/C11.H2'1	1.622
115: B/C11.H2'2	2.209
116: B/C11.H3'	4.399
117: B/C11.H6	7.385
118: B/C11.H5	5.515
119: B/C11.H41	?
120: B/C11.H42	?
121: B/C12.H5'1	3.840
122: B/C12.H5'2	3.763
123: B/C12.H4'	4.297
124: B/C12.H1'	5.432
125: B/C12.H2'1	2.198
126: B/C12.H2'2	1.996
127: B/C12.H3'	4.652
128: B/C12.H6	7.346
129: B/C12.H5	5.442
130: B/C12.H41	?
131: B/C12.H42	?
132: B/G13.H5'1	3.941
133: B/G13.H5'2	3.805
134: B/G13.H4'	4.222
135: B/G13.H1'	5.724
136: B/G13.H2'1	2.406
137: B/G13.H2'2	2.549
138: B/G13.H3'	4.851
139: B/G13.H8	7.766
140: B/G13.H1	?

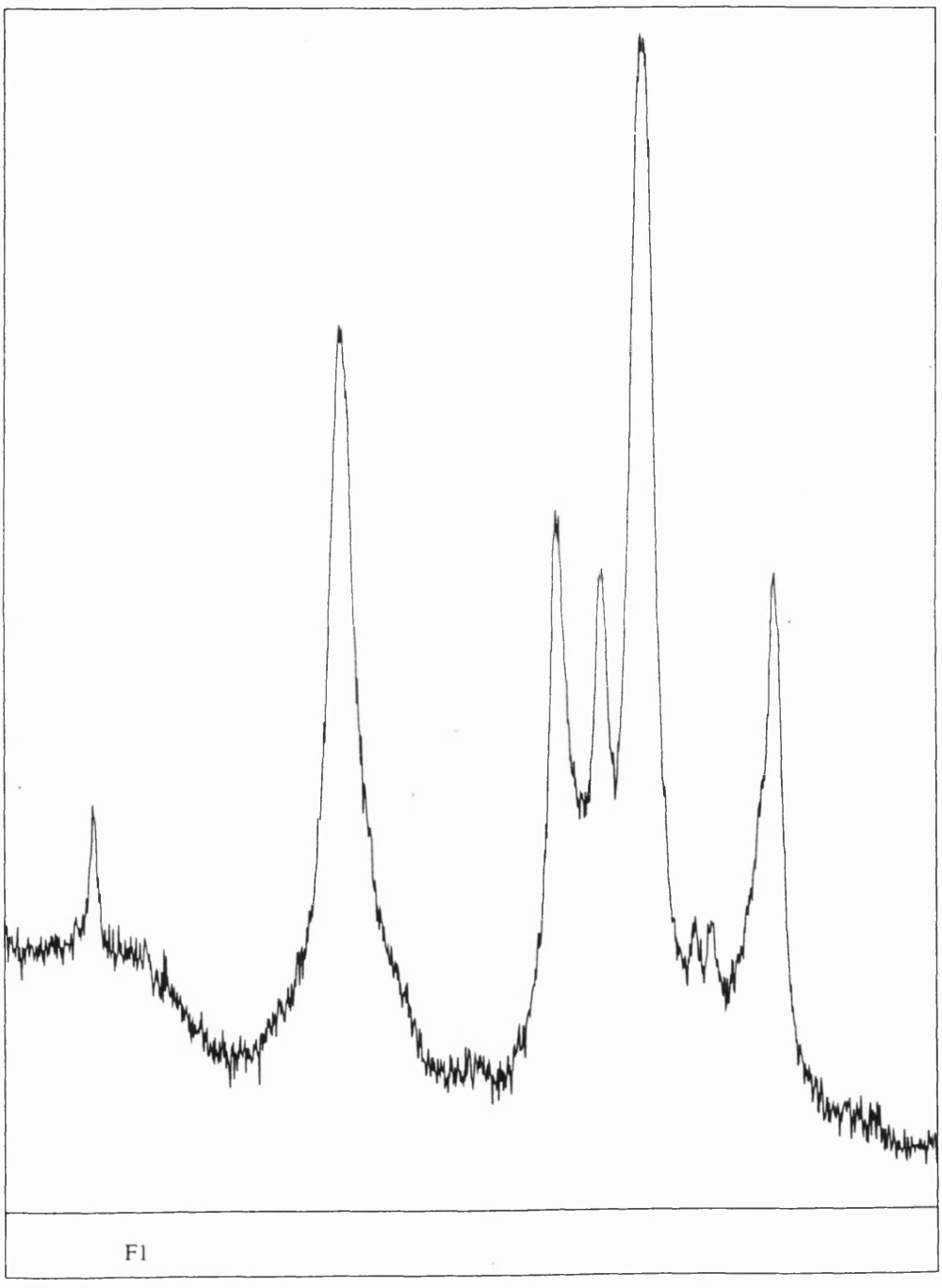
	1: SHIFT
141: B/G13.H21	?
142: B/G13.H22	?
143: B/T14.H5'1	4.085
144: B/T14.H5'2	4.010
145: B/T14.H4'	4.240
146: B/T14.H1'	6.142
147: B/T14.H2'1	2.206
148: B/T14.H2'2	2.546
149: B/T14.H3'	4.827
150: B/T14.H6	7.454
151: B/T14.H51	1.617
152: B/T14.H52	1.617
153: B/T14.H53	1.617
154: B/T14.H3	?
155: B/T15.H5'1	4.249
156: B/T15.H5'2	4.016
157: B/T15.H4'	2.566
158: B/T15.H1'	6.061
159: B/T15.H2'1	1.383
160: B/T15.H2'2	1.614
161: B/T15.H3'	4.304
162: B/T15.H6	7.417
163: B/T15.H51	1.466
164: B/T15.H52	1.466
165: B/T15.H53	1.466
166: B/T15.H3	?
167: B/T16.H5'1	3.444
168: B/T16.H5'2	2.908
169: B/T16.H4'	3.973
170: B/T16.H1'	5.708
171: B/T16.H2'1	2.190
172: B/T16.H2'2	1.460
173: B/T16.H3'	4.389
174: B/T16.H6	7.179
175: B/T16.H51	1.584

	1: SHIFT
176: B/T16.H52	1.584
177: B/T16.H53	1.584
178: B/T16.H3	?
179: B/T17.H5'1	3.729
180: B/T17.H5'2	3.442
181: B/T17.H4'	3.119
182: B/T17.H1'	5.066
183: B/T17.H2'1	1.720
184: B/T17.H2'2	1.837
185: B/T17.H3'	4.304
186: B/T17.H6	7.016
187: B/T17.H51	1.378
188: B/T17.H52	1.378
189: B/T17.H53	1.378
190: B/T17.H3	?
191: B/T18.H5'1	3.606
192: B/T18.H5'2	3.235
193: B/T18.H4'	3.908
194: B/T18.H1'	5.438
195: B/T18.H2'1	2.181
196: B/T18.H2'2	1.396
197: B/T18.H3'	4.404
198: B/T18.H6	6.947
199: B/T18.H51	1.347
200: B/T18.H52	1.347
201: B/T18.H53	1.347
202: B/T18.H3	?
203: B/C19.H5'1	3.306
204: B/C19.H5'2	2.892
205: B/C19.H4'	2.156
206: B/C19.H1'	5.163
207: B/C19.H2'1	1.914
208: B/C19.H2'2	1.489
209: B/C19.H3'	4.519
210: B/C19.H6	7.032

	1: SHIFT
211: B/C19.H5	5.356
212: B/C19.H41	?
213: B/C19.H42	?
214: B/G20.H5'1	3.928
215: B/G20.H5'2	3.894
216: B/G20.H4'	4.047
217: B/G20.H1'	5.887
218: B/G20.H2'1	2.432
219: B/G20.H2'2	2.198
220: B/G20.H3'	4.516
221: B/G20.H8	7.658
222: B/G20.H1	?
223: B/G20.H21	?
224: B/G20.H22	?
225: A/OH1.H3	7.901
226: A/OH1.H6	7.031
227: A/OH1.M3	2.896
228: A/OH1.M2	2.896
229: A/OH1.M1	2.896
230: A/OH1.H8A2	4.300
231: A/OH1.H8A	3.730
232: A/OH1.H8B	5.153
233: A/OH1.H1A	3.715
234: A/OH1.H1B	4.813
235: A/OH1.H7'1	7.623
236: A/OH1.H6'1	8.110
237: A/OH1.H4'1	7.755
238: A/OH1.H3'1	7.450
239: A/OH1.H4'2	7.822
240: A/OH1.H5'2	7.475
241: A/OH1.H6'2	7.308
242: A/OH1.H7'2	8.000
243: A/OH1.H3'2	7.903

APPENDIX XX

Figure 178 Expansion of the methyl region of the 1D NMR spectrum of 5′-(GTGATGAG)·5′-(CACTACTC) DNA duplex after the resction with DSB-120.



Bibliography

Abraham, R. J., Fisher, J., & Loftus, P. (1988) *Introduction to NMR Spectroscopy*, Second Edn. John Wiley & Sons, Chichester, UK.

Abu-Daya, A., Brown, P. M., & Fox, K. R. (1995) *Nucleic Acids Res.* 23, 17, 3385-3392.

Acton, E. M., Tong, G. L., Mosher, C. W., & Wolgemuth, R. L. (1984) *J. Med. Chem.* 27, 638-645.

Ades, S. E., & Sauer, R. T. (1994) *Biochemistry* 33, 9187-9194.

Animati, F., Arcamone, F. M., Conte, M. R., Felicetti, P., Galeone, A., Lombardi, P., Mayol, L., Paloma, L. G., & Rossi, C. (1995) *J. Med. Chem.* 38, 11400-1149.

Applied Biosystems *Model 381A DNA Synthesizer manual.*, Sec.2.1-20

Arcamone, F., Penco, S., Orezzi, P., Nicolella, V., & Pirelli, A. (1964) *Nature* 203, 1064-1065.

Arcamone, F., Cassinelli, G., Fantini, G., Grein, A., Orezzi, P., Pol, C., & Spalla, C. (1969) *Biotechnol. Bioeng.* 11, 1101.

Arima, K., Kohsaka, M., Tamura, G., Imanaka, H., & Sakai, H. (1972) *J. Antibiot.* 24, 437-444.

Arora, S. K. (1981) *J. Antibiot.* 34, 462-464.

Auffinger, P., Louise-May, S., & Westhof, E. (1995) *J. Am. Chem. Soc.* 117, 6720-6726.

Baguley, B. C. (1982) *Molec. Cell. Biochem.* 43, 167-181.

Baraldi, P. G., Gacciari, B., Romagnoli, R., Spalluto, G., Gambari, R., Bianchi, N., Passadore, M., Ambrosino, P., Mongelli, N., Cozzi, P., & Geroni, C. (1997) *Anti-Cancer Drug Design* 12, 555-576.

Barkley, M. D., Cheatham, S., Thurston, D. E., Hurley, L. H. (1986) *Biochemistry* 25, 3021-3031.

Begleiter, A., & Johnston, J. B. (1985) *Biochem. Biophys. Res. Commun.* 131, 1, 336-338.

Berman, H. M., Neidle, S., Zimmer, C. H., & Thrum, H. (1979) *Biochim. Biophys. Acta.* 561, 124-131.

Bhuyan, B. K., & Dietz, A. (1965) *Antimicrob. Ag. Chemother.* 836-844.

Bhuyan, B. K., Newell, K. A., Adams, E. G., Crampton, S. L., & Van Hoff, D. D. (1981) *Proc. Am. Assoc. Cancer Res.* 22, 224.

Bhuyan, B. K., Newell, K. A., Crampton, S. L., & Von Hoff, D. D. (1982) *Cancer Res.* 42, 3532-3537.

Bhuyan, B. K., Smith, K. S., Adams, E. G., Petzold, G. L., McGovren, J. P. (1992) *Cancer Res.* 52, 5687-5692.

Bio-Rad., *DNA Grade Bio-Gel HTP Hydroxylapatite Instruction manual.*

Blake, P. R., & Summers, M. F. (1990) *J. Magn. Reson.* 86, 622-624.

Blanco. M. (1991) *J. Comp. Chem.* 12, 2, 237-247.

Blasko, A., & Bruice, T. C. (1993) *Proc. Natl. Acad. Sci. USA* 90, 10018-10022.

Boger, D. L., Coleman, R. S., Invergo, B. J., Sakya, S. M. Ishizaki, T., Munk, S. A., Zarrinmayeh, H., Kitos, P. A., & Collins Thompson S. (1990) *J. Am. Chem. Soc.* 112, 4623-4632.

Boger, D. L., & Ishizaki, T. (1990) *Tetrahedron Lett.* 31, 6, 793-796.

Boger, D. L., & Johnson, D. S. (1995) *J. Am. Chem. Soc.* 117, 1443-1444.

Boger, D. L., & Mésini, P. (1995) *J. Am. Chem. Soc.* 117, 11647-11655.

Boger, D. L., Munk, S. A., & Ishizaki, T. (1991b) *J. Am. Chem. Soc.* 113, 2779-2780.

Boger, D. L., Munk, S. A. & Zarrinmayeh, H.(1991a) *J. Am. Chem. Soc.* 113, 3980-3983.

Boger, D. L., Yun, W., & Han, N. (1995) *Bioorg. Med. Chem.* 3, 11, 1429-1453.

Bontemps, J., Hovssier, C., & Fredericq, E. (1975) *Nucleic Acids Res.* 2, 971-984.

Bose, D. S., Thompson, A. S., Ching, J., Hartley, J. A., Berardini, M. J., Jenkins, T. C., Neidle, S., Hurley, L. H., & Thurston, D. E. (1992) *J. Am. Chem. Soc.* 114, 4939-4941.

Boyd, D. B., & Lipkowitz, K. B. (1982) *J. Chem. Educ.* 59, 4, 269-274.

Boyd, F. L., Stewart, D., Remers, W. A., Barkley, M. D., & Hurley, L. H. (1990) *Biochemistry* 29, 2387-2403.

Braithwaite, A. W., & Baguley, B. C. (1980) *Biochemistry* 19, 1101-1106.

Brown, D. G., Sanderson, M. R., Garman, E., & Neidle, S. (1992) *J. Mol. Biol.* 226, 481-490.

Brown, D. G., Sanderson, M. R., Skelly, J. V., Jenkins, T. C., Brown, T., Garman, E., Stuart, D. I., & Neidle, S. (1990) *Embo. J.* 9, 1329-1334.

Burris, H. A., Dieras, V. C., Tunca, M., Earhart, R. H., Eckardt, J. R., Rodriguez, G. I., Shaffer, D. S., Fields, S. M., Campbell, E., Schaaf, L., Kasunic, D., & Von Hoff, D. D. (1997) *Anti-Cancer Drugs* 8, 588-596.

Cain, B. F., Atwell, G. J., & Seelyé, R. N. (1969) *J. Med. Chem.* 12, 199-206.

Cameron, L., & Thompson, A. S. (1999) *Biochemistry* In Press.

Capranico, G., Soranzo, C., & Zunino, F. (1986) *Cancer Res.* 46, 5499-5503.

Cargill, C., Bachmann, E., & Zbinden, G. (1974) *J. Natl. Cancer Inst.* 53, 481-486.

Chaires, J. B., Fox, K. R., Herrera, J. E., Britt, M., & Waring, M. J. (1987) *Biochemistry* 26, 8227-8236.

Chary, K. V. R., Hosur, R. V., Govil, G., Zu-kun, T., & Miles, H. T. (1987) *Biochemistry* 26, 1315-1322.

Chazin, W. J., Wüthrich, K., Hyberts, S., Rance, M., Denny, W. A., & Leupin. W. (1986) *J. Mol. Biol.* 190, 439-453.

Chen, S.-M., Leupin, W., Rance, M., & Chazin, W. J. (1992) *Biochemistry* 31, 4406-4413.

Chen, A. Y., & Liu, L. F. (1994) *Annu. Rev. Pharmacol. Toxicol.* 34, 191-218.

Chen, Y.-L., & Lown, J. W. (1994) *J. Am. Chem. Soc.* 116, 16, 6995-7005.

Chen, X., Ramakrishnan, B., & Sundaralingam, M. (1997) *J. Mol. Biol.* 267, 1157-1170.

Chiang, S.-Y., Welch, J., Rausher, F. J. III., & Beerman, T. A. (1994) *Biochemistry* 33, 70333-7040.

Chidester, C. G., Krueger, W. C., Mizsak, S. A., Duchamp, D. J., & Martin, D. G. (1981) *J. Am. Chem. Soc.* 103, 7629-7635.

Clark, G. R., Boykin, D. W., Czarny A., & Neidle, S. (1997) *Nucleic Acids Res.* 25, 8, 1510-1515.

Clark, G. R., Squire, C. J., Gray, E. J., Leupin, W., & Neidle, S. (1996) *Nucleic Acids Res.* 24, 24, 4882-4889.

Cobuzzi, R. J., Burhans, W. C., & Beerman, T. A. (1996) *J. Biol. Chem.* 271, 33, 19852-19859.

Coll, M., Aymami, J., van der Marel, G. A., van Boom, J. H., Rich, A., & Wang, A. H.-J. (1989) *Biochemistry* 28, 310-320.

Coll, M., Frederick, C. A., Wang, A. H.-J., & Rich, A. (1987) *Proc. Natl. Acad. Sci. USA* 84, 8385-8389.

Conte, M. R., Jenkins, T. C., & Lane, A. N. (1995) *Eur. J. Biochem.* 229, 433-444.

Cristofanilli, M., Bryan, W. J., Miller, L. L., Chang, A. Y,-C., Gradishar, W. J., Kufe, D. W., & Hortobagyi, G. N. (1998) *Anti-Cancer Drugs* 9, 779-782.

Cullinane, C., Cutts, S. M., van Rosmalen, A., Phillips, D. R. (1994) *Nucleic Acids Res.* 22, 12, 2296-2303.

Cullinane, C., & Phillips, D. R. (1991) *FEBS* 293, 1, 2, 195-198.

Cullinane, C., & Phillips, D. R. (1992) *Biochemistry* 31, 40, 9513-9519.

Cullinane, C., & Phillips, D. R. (1993) *Nucleic Acids Res.* 21, 8, 1857-1862.

Cullinane, C., & Phillips, D. R. (1994) *Biochemistry* 33, 6207-6212.

Dano, K. (1971) *Cancer Chemother. Rept.* 55, 133.

Denny, W. A., Atwell, G. J., Baguley, B. C., & Cain, B. F. (1979) *J. Med. Chem.* 22, 134-150.

Dervan, P. B. (1986) *Science* 232, 464-471.

Dickinson, L. A., Guliza, R. J., Trauger, J.W., Baird, E.E., Mosier, D. E., Gottesfeld, J. M., & Dervan, P. B. (1998) *Proc. Natl. Acad. Sci. USA.* 95, 12890-12895.

Di Marco, A., Gaetani, M., Orezzi, P., Scarpinato, B., Silvestrini, R., Soldati, M., Dasdia, T., & Valentini, L. (1964) *Nature* 201, 706.

Duncan, R. (1992) *Anticancer Drugs* 3, 175.

Dwyer, T. J., Geierstanger, B. H., Bathini, Y., Lown, J. W., & Wemmer, D. E. (1992) *J. Am. Chem. Soc.* 114, 15, 5911-5919.

Embrey, K. J., Searle, M. S., & Craik D. J. (1993) *Eur. J. Biochem.* 211, 437-447.

Ettorre, A., Cirilli, M., & Ughetto, G. (1998) *Eur. J. Biochem.* 258, 2, 350-354.

Evans, J. N. S. (1995) *Biomolecular NMR Spectroscopy* 115-144, Oxford University Press, Oxford, UK.

Fede, A., Labhardt, A., Bannwarth, W., & Leupin, W. (1991) *Biochemistry* 30, 11377-11388.

Feigon, J., Denny, W. A., Leupin, W., & Kearns, D. R. (1983a) *Biochemistry* 22, 5930-5942.

Feigon. J., Leupin. W., Denny, W. A., & Kearns, D. R. (1983b) *Biochemistry* 22, 5943-5951.

Feigon, J., Wright, J. M., Leupin, W., Denny, W. A. & Kearns, D. R. (1982) *J. Am. Chem. Soc.* 104, 5540-5541.

Filippini, C., Bisiach, M., Tagliabue, G., D'Incalci, M., Ubezio, P. (1997) *Int. J. Cancer* 72, 801-809.

Finlay, A. C., Hochstein, F. A., Sobin, B. A., & Murphy, F. X. (1951) *J. Am. Chem. Soc.*, 73, 341-343.

Fisher, J. F., & Aristoff, P. A. (1988) *Prog. Drug Res.* 32, 411-498.

Frenkiel, T. A. (1993) in *NMR of Macromolecules A Practical Approach* 35-70, Roberts, G. C. K. (ed.) IRL Press, Oxford, UK.

Gait, M J. (ed.) (1984) *Oligonucleotide Synthesis – A Practical Approach*, IRL Press, Oxford, UK.

Gao, Y. G., Sriram, M., Denny. W. A., & Wang, A. H.-J. (1993) *Biochemistry* 32, 9639-9648.

Gao, Y.-G., Liaw, Y.-C., Robinson, H., & Wang, A. H.-J. (1990) *Biochemistry* 29, 10307-10316.

Garbesi, A., Tondelli, L., Conte, M. R., Galeone, A., Paloma, L. G., & Mayol, L. (1996) *Gazzetta Chimica Italiana* 126, 799-803.

Gasteiger, J., & Marsili, M. (1980) *Tetrahedron* 36, 3219-3288.

Gause, G. F., & Dudnik, Y. V. (1971) *Progr. Mol. Subcell. Biol.* 2, 33-39.

Gause, G. F. (1975) in *Antibiotic III* (Corcoran, J. W., and Hahn F. E. eds.) 269-273, Academic Press, New York.

Gause, G. F., Preobrazhenskaya, T. P., Invanitskaya, L. P., & Sveshnikova, M. A. (1969) *Antibiotiki* 14, 963-669.

Goodsell, D. S., Kaczor-Grzeskowiak, M., & Dickerson, R. E. (1994) *J. Mol. Biol.* 239, 79-96.

Gorbunoff, M. J. (1984) *Anal. Biochem.* 136, 425-432.

Graves, D. E., Pattaroni, C., Krishnan, B. S., Ostrander, J. M., Hurley, L. H., & Krugh, T. R. (1984) *J. Biol. Chem.*259, 13, 8202-8209.

Graves, D. E., Stone, M. P., Krugh, T. R. (1985) *Biochemistry* 24, 7573-7581.

Greenidge, P. A., Jenkins, T. C., & Neidle, S. (1993) *Molec. Pharmacol.* 43, 982-988.

Gregson, S. J., Howard, P. W., Jenkins, T. C., Kelland, L. R., & Thurston, D.E. (1999) *Chem. Commun.* 9, 797-798.

Gunz, D., & Naegeli, H. (1996) *Biochem. Pharmacol.* 52, 447-453.

Gurskaya, G. V., Grokhovsky, S. L., Zhuze, A. L., & Gottikh, B. P. (1979) *Biochim. Biophys. Acta. 563, 336-342.*

Hahn, F. E. (1975) in *'Antibiotics III'* 79-100 (eds Corcoran, J. W., & Hahn, F. E.) Springer-Verlag, Berlin – Heidleberg – New York,.

Hanka, L. J., Dietz. A., Gerpheide, S. A., Kuentzel, S. L., & Martin, D. G. (1978) *J. Antibiot.* 31, 1211-1217.

Hannan, M. A., & Hurley, L. H. (1978) *Cancer Res.* 38, 2795-2799.

Hansen, M., & Hurley, L. H., (1995) *J. Am. Chem. Soc.* 117, 2421-2429.

Hansen, M., Yun, S., & Hurley, L (1995) *Chem. Biol.* 2, 4, 229-240.

Hare, D. R., Wemmer, D. E., Chou, S.-H., & Drobny, G. (1983) *J. Mol. Biol.* 171, 319-336.

Herman, D. E., Baird, E. E., & Dervan, P. B. (1998) *J. Am. Chem. Soc.* 120. 7 1382-1391.

Hertzberg, R. P., Hecht, S. M., Reynolds, V. L., Molineux, I. J., & Hurley, L. H. (1986) *Biochemistry* 25, 1249-1258.

Horwitz, S. B., & Grollman, A. P. (1968) *Antimicrob. Agents Chemother.* 21-24.

Hu, S., Weisz, K., James, T. L., Shafer, R. H. (1992) *Eur. J. Biochem.* 204, 31-38.

Hurley, L. H. (1977) *J. Antibiot.* 30, 349-370.

Hurley, L. H., Chandler, C., Garner, T., Petrusek, P. L., & Zimmer, S.(1979) *J. Biol. Chem.* 254, 605-608.

Hurley, L. H., & Draves, P. H. (1993) in *Molecular Aspects of Anticancer Drug-DNA Interactions Volume 1,* Neidle, S., & Waring, M. (eds.) The Macmillan Press LTD, London, UK.

Hurley, L. H., Gairola, C., & Zmijewski, M. (1977) *Biochim. Biophys. Acta* 475, 521-535.

Hurley, L. H., Lee, C.-S., McGovren, J. P., Warpehoski, M. A., Mitchell, M. A., Kelly, R. C., & Aristoff, P. A. (1988a) *Biochemistry* 27, 3886-3892.

Hurley, L. H., & Needham-VanDevanter, D. R. (1986) *Acc. Chem. Res.* 19, 230-237.

Hurley, L. H., Needham-VanDevanter, D. R., & Lee, C.-S. (1987) *Biochemistry* 84, 6412-6416.

Hurley, L. H., Reck. T., Thurston, D. E., Langley, D. R., Holden, K. G., Hertzberg, R. P., Hoover, J. R. E., Gallagher, G. Jr., Faucette, L. F., Mong, S.-M., & Johnson, R. K. (1988b) *Chem. Res. Toxicol.* 1, 258-268.

Hurley, L. H., Reynolds, V. L., Swenson. D. H., Petzold, G. L., & Scahill, T. A. (1984) *Science* 226, 843-844.

Hurley, L. H., Rokem, J. S., & Petrusek. R. L. (1980) *Biochem. Pharmacol.* 29, 1307-1310.

Hurley, L.H., & Thurston, D. E. (1984) *Pharm. Res.* 1, 52-59.

Hurley, L. H., Warpehoski, M. A., Lee, C.-S., McGovren, J. P., Scahill, T. A., Kelly, R. C., Mitchell, M. A., Wicnienski, N. A., Gebhard, I., Johnson, P. D., & Bradford, V. S. (1990) *J. Am. Chem. Soc.* 112, 4633-4649.

Jacobson, M. K., Twehous, D., & Hurley, L. H. (1986) *Biochemistry* 25, 5929-5932.

Jenkins, T. C., Hurley, L. H., Neidle, S., & Thurston, D. E. (1994) *J. Med. Chem.* 37, 4529-4537.

Jenkins, T. C., & Lane, A. N. (1997) *Biochim. Biophys. Acta* 1350, 189-204.

Jenkins, T. C., Lane, A. N., Neidle, S., & Brown, D. G. (1993) *Eur. J. Biochem.* 213, 1175-1184.

Jesson, M. I., Johnston, J. B., Robotham, E., & Begleiter, A. (1989) *Cancer Res.* 49, 7031-7036.

Kaplan, D. J. (1982) *Biochem. Biophys. Res. Commun.* 109 639-648.

Karplus, M., & Petsko, G. A. (1990) *Nature.* 347, 631-639.

Katahira, M., Sugeta, H., Kyogoku, Y., Fujii, S., Fujisawa, R., & Tomita, K. (1988) *Nucleic Acid Res.* 16, 17, 8619-8632.

Kielkopf, C. L., Baird, E. E., Dervan, P. D., & Rees, D. C. (1998) *Nat. Struct. Biol.* 5, 104-109.

Kintanar, A., Klevit, R. E., & Reid, B. R. (1987) *Nucleic Acids Res.* 15, 14, 5845-5861.

Klevit, R. E., Wemmer, D. E., & Reid, B. R. (1986) *Biochemistry* 25, 3296-3303.

Kopka, M. L., Goodsell, D. S., Baikalov, I., Grzeskowiak, K., Cascio, D., & Dickerson, R. E. (1994) *Biochemistry* 33, 13593-13610.

Kopka, M. L., & Larsen, T. A. (1992) in *Nucleic Acid Targeted Drug Design,* 303-374, Propst, C L., & Perun, T. J. (eds.), Marcel Dekker, Inc., New York, USA.

Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985) *J. Mol. Biol.* 183, 553-563.

Krowicki, K., Balzarini, J., De Clercq, E., Newman, R. A., & Lown, J. W. (1988) *J. Med. Chem.* 31, 341-345.

Krugh, T. R., Graves, D. E., & Stone, M. P. (1989) *Biochemistry* 28, 9988-9994.

Lane, A. N., Jenkins, T. C., Brown, T., Neidle, S. (1991) *Biochemistry* 30, 1372-1385.

Larsen, T., Goodsell, D. S., Cascio, D., Grzeskowiak, K., & Dickerson, R. E. (1989) *J. Biomol. Struct. Dyn.* 7, 477-491.

Lee, C.-S., & Gibson, N. W. (1991) *Cancer Res.* 51, 6586-6591.

Lee, C.-S., & Gibson, N. W. (1993) *Biochemistry* 32, 9108-9114.

Lee, C. S., & Hurley, L. H. (1986) *Proc. Am. Assoc. Cancer Res.* 27, 243.

Lee, C.-S., Myung, P.-K., & Gibson, N. W. (1996) *Arch. Pharm. Res.* 19, 3, 191-196.

Lee, C.-S., Pfeifer, G. P., Gibson, N. W. (1994) *Biochemistry* 33, 6024-6030.

Lee, C. S., Sun, D., Kizu, R., & Hurley, L. H. (1991) *Chem. Res. Toxicol.* 4, 203-213.

Leupin, W., Chazin, W. J., Hyberts, S., Denny, W. A., & Wüthrich, K. (1986) *Biochemistry* 25, 5902-5910.

Li, L. H., Dekoning, T. F., Kelly, R. C., Krueger, W. C., McGovren, J. P., Padbury, G. E., Petzold, G. L., Wallace, T. L., Ouding, R. J., Prairie, M. D., & Gebhard, I. (1992) *Cancer Res.* 52, 4904-4913.

Li, L. H., Kuentzel, S. L., Murch, L. L., Pschigoda, L. M., & Krueger, W. C. (1979) *Cancer Res.* 39, 4816-4822.

Li, L. H., Swenson, D. H., Schpok, S. L. F., Kuentzel, S. L., Dayton, B. D., & Krueger, W. C. (1982) *Cancer Res.* 42, 999-1004.

Lin, C. H., Beale, J. M., & Hurley, L. H. (1991) *Biochemistry* 30, 3597-3602.

Lin, C. H. & Hurley, L. H. (1990) *Biochemistry* 29, 9503-9507.

Loewe, H., & Urbanietz, J. (1968) in *Abstracts and Reviews of Eighth International Congress on Tropical Medicine and Malaria,* 139, Teheran, Iran.

Martin, D. G., Hanka, L. H., Neil, G. L. (1978) *Proc. Am. Assoc. Cancer Res.* 19, 99.

Maruyama, I. N., Suzuki, H., & Tanaka, N. (1978) *J. Antibiot.* 31 761-768.

Mazzini, S., Mondelli, R., & Ragg, E. (1998) *J. Chem. Soc.* Perkin Trans. 2, 1983-1991.

McGovren J. P., Clarke, G. L., Pratt, E. A., DeKoning, T. F. (1984) *J. Antibiot.* 37, 1, 63-70.

McHugh, M. M., Woynarowski, J. M., Mitchell, M. A., Gawron, L. S., Weiland, K. L., & Beerman, T. A. (1994) *Biochemistry* 33, 9158-9168.

Metzler, W. J., Wang, C., Kitchen, D. B., Levy, R. M., & Pardi, A. (1990) *J. Mol. Biol.* 214, 711-736.

Mitchell, M. A., Kelly, R. C., Winnienski, N. A., Hatzenbuhler, N. T., & Williams, M. G. (1991) *J. Am. Chem. Soc.* 113, 8994-8995.

Morrison, R. T., & Boyd, R. N. (1992) *Organic Chemistry,* Prentice Hall International, Inc., London, UK.

Mostad, A., Rømming, C., & Storm, B. (1978) *Acta Scand. Ser. B.* 32, 639-645.

Mountzouris, J. A., Wang, J.-J., Thurston, D., & Hurley, L. H. (1994) *J. Med. Chem.* 37, 3132-3140.

Mrksich, M., Parks, M. E., & Dervan, P. B. (1994) *J. Am. Chem. Soc.* 7983-7988.

Neil, G. L., Clarke, G. L., & McGovren, J. P. (1981) *Proc. Am. Assoc. Cancer Res.* 22, 244.

Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature* 330, 221-226.

Newton, B. A., (1975) *Antibiotics III, Mechanism of Action Antimicrobial and Antitumor Agents* (eds. Corcoran, J. W., & Hahn, F. E.) 34-47 Springer, Verlag, Berlin.

Nguyen, H. N., Sevin, B. U., & Averette, H. (1992) *Cancer Chemother. Phaarmacol.* 30, 37-42.

Nunn, C. M., & Neidle, S. (1995) *J. Med. Chem.* 38, 2317-2325.

Odefey, C., Westendorf, J., Dieckmann, T., & Oschkinat, H. (1992) *Chem.-Biol. Interact.* 85, 117-126.

Onetto, N., LoBugli, A., Brookman, M., Gilewsky, T., Dougan, M., Healey, D., Hellstrom, K. E., Trial, P., Siegall, C., Birkhoter, M., & Canetta, R. (1995) *Proceedings, 2nd EORTC-EDDM,* June 21-24, Corfu, Greece, p59.

Parkinson, J. A., Barber, J., Douglas, K. T., Rosamond, J., & Sharples, D. (1990) *Biochemistry* 29 10181-10190.

Parkinson, J. A., Ebrahimi, S. E., McKie, J. H., & Douglas, K. T. (1994) *Biochemistry* 33, 8442-8452.

Parks, M. E., Baird, E. E., & Dervan, P. B., (1996) *J. Am Chem. Soc.* 118, 26, 6153-6159.

Parolin, C., Montecucco, A., Chiarocchi, G., Pedrali-Noy, G., Valisena, S., Palumbo, M., & Palù, G. (1990) *FEMS Microbiol. Lett.* 68, 341-346.

Patel. D. J. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6424-6428.

Patel, D. J., Shapiro, L., & Hare, D. (1986) *J. Biol. Chem.* 261, 3, 1223-1229.

Pearl, L. H., Skelly, J. V., Hudson, B. D., & Neidle, S. (1987) *Nucleic Acids Res.* 15, 8, 3469-3477.

Pelton, J. G., & Wemmer, D. E. (1988) *Biochemistry* 27, 8088-8096.

Pelton, J. G., & Wemmer, D. E. (1989) *Proc. Natl. Acad. Sci. USA* 86, 5723-5727.

Pelton, J. G., & Wemmer, D. E. (1990) *J. Am. Chem. Soc.* 112, 1393-1399.

Petrusek, R. L., Anderson, G. L., Garner, T. F., Fannin, Q. L., Kaplan, D. J., Zimmer, S. G., & Hurley, L. H. (1981) *Biochemistry* 20, 1111-1119.

Pierce, J. R., Nazimiec, M., & Tang, M.-S. (1993) *Biochemistry* 32, 7069-7078.

Pilch, D. S., Kirolos, M. A., Liu, X., Plum, G. E., & Breslauer, K. J. (1995) *Biochemistry* 34, 9962-9976.

Portugal, J. (1994) *FEBS Lett.* 344, 136-138.

Portugal, J., & Waring, M. J. (1987) *Eur. J. Biochem.* 167, 281-289.

Powers R., & Gorenstein D. G. (1990) *Biochemistry* 29, 9994-10008.

Primrose, W. U. (1993) in *NMR of Macromolecules A Practical Approach,* 7-34, Roberts, G. C. K. (ed.) IRL Press, Oxford, UK.

Purcell, W. P., Singer, J. A. (1967) *J. Chem. Eng. Data* 12, 2, 235-246.

Puvvada, M. S., Forrow, S. A., Hartley, J. A., Stephenson, P., Gibson, I., Jenkins, T. C., & Thurston, D. E. (1997) *Biochemistry* 36, 9, 2478-2484.

Raether, W., & Lammler, G., (1971) *Annals. Tropical Med. Parasit.* 65, 107-115.

Ragg, E., Mazzini, S., Bortolini, R., Mongelli, N., & D'Alessio, R. (1998) *J. Chem. Soc. Perkin Trans.* 2, 149-158.

Remers, W. A., Mabilia, M., & Hopfinger, A. J. (1986) *J. Med. Chem.* 29, 2492-2503.

Remin, M., & Shugar, D. (1972) *Biochem. Biophys. Res. Commun.* 48, 3, 636-642.

Reynolds, V. L., Molineux, I. J., Kaplan, D. J., Swenson, D. H., & Hurley L. H. (1985) *Biochemistry* 24, 6228-6237.

Rhodes, D., & Klug, A. (1981) *Nature* 292, 378-380.

Ridler, P. J., & Jennings, B. R. (1980) *Int. J. Biol. Macromol. 2, 313-317.*

Robinson, H., Liaw, Y. C., van der Marel, G. A., van Boom, J. H., & Wang, A. H. -J. (1990) *Nucleic Acid Res.* 18, 4851-4858.

Robinson, H., Yang, D., & Wang, A. H.-J. (1994) *Gene* 149, 179-188.

Rydzeqski, J. M., Leupin, W., & Chazin, W. (1996) *Nucleic Acids Res.* 24, 7, 1287-1293.

Scahill, T. A., Jensen, R. M., Swenson, D. H., Hatzenbuhler, N. T., Petzold, G., Wierenga, W., & Brahme, N. D. (1990) *Biochemistry* 29, 2852-2860.

Scheek, R. M., Boelens, R., Russo, N., van Boom, J. H., Kaptein, R. (1984) *Biochemistry* 23, 1371-1376.

Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., & van Boom, J. H. (1983) *J. Am. Chem. Soc.* 105, 2914-2916.

Seaman, F. C., Chu, J., & Hurley, L. (1996) *J. Am. Chem. Soc.* 118, 5383-5395.

Seaman, F. C., & Hurley, L. H. (1993) *Biochemistry* 32, 12577-12585.

Seaman, F. C., & Hurley, L. (1996) *J. Am. Chem. Soc.* 118, 10052-10064.

Searle, M. S. (1993) *Prog. NMR Spec.* 25, 403-480.

Searle, M. S., & Bicknell, W. (1992) *Eur. J. Biochem. 205,* 45-58.

Searle, M. S., & Embrey, K. J. (1990) *Nucleic Acids Res.* 18, 13, 3753-3762.

Searle, M. S., Hall, J. G., Denny, W. A., & Wakelin, L. P. G. (1988) *Biochemistry* 27, 4340-4349.

Sen, D., & Crothers, D. (1986) *Biochemistry* 25, 1503-1509.

Shamdas, G. J., Alberts, D. S., Modiano, M., Wiggins, C., Power, J., Kasunic, D. A., Elfring, G. L., & Earhart, R. H. (1994) *Anti-Cancer Drugs* 5, 10-14.

Sklenar, V., & Bax, A. (1987) *J. Magn. Reson.* 74, 469-479.

Smith, C., K., Brannigan, J. A., & Moore, M. H. (1996) *J. Mol. Biol.* 263, 237-258.

Sorio, R., Wanders, J., Comella, G., Dittrich, C., Droz, J. P., Morant, R.,Cognetti, F., & Hanauske, A. –R. (1998) *Annals. Oncol.* 9, 674.

Storl, K., Storl. J., Zimmer, C. H., & Lown, J. W. (1993) *FEBS Lett.* 317, 157-162.

Stryer, L. (1988) *Biochemistry,* W.H. Freeman and company, New York, USA

Sun, D., & Hurley, L. H. (1992a) *Biochemistry* 31, 2822-2829.

Sun, D., & Hurley, L. H. (1992b) *J. Med. Chem.* 35, 1773-1782.

Sun, D., & Hurley, L. H. (1992) *Anti-Cancer Drug Design* 7, 15-36.

Sun, D., & Hurley, L. H. (1993) *J. Am. Chem. Soc.* 115, 5925-5933.

Surovaya, A. N., Burckhardt, G., Grokhovsky, S. L., BirchHirschfeld, E., Gursky, G. V., & Zimmer, C. (1997) *J. Biomol. Struct. Dyn.* 14, 5, 595-606.

Swalley, S. E., Baird, E. E., & Dervan, P. B. (1997) *J. Am. Chem. Soc.* 119, 30, 6953-6961.

Swalley, S. E., Baird, E. E., & Dervan, P. B. (1999) *J. Am. Chem. Soc.* 121, 6, 1113-1120.

Swenson, D. H., Li, L. H., Hurley, L. H., Rokem, J. R., Petzold, G. L., Dayton, B. D., Wallace, T. L., Lin, A. H., & Krueger, W. C. (1982) *Cancer Res. 42,* 2821-2828.

SYBYL® Version 6.2 Manual, (1995) Tripos. Inc., St. Louis, Missouri, USA.

Takeuchi, T., Miyamoto, M., Ishizuka, M., Naganawa, H., Kondo, S., Hamada, M., & Umezawa, H. (1976) *J. Antibiot.* 29, 93-96.

Tanious, F. A., Ding, D., Patrick, D. A., Tidwell, R. R., & Wilson, W. D. (1997) *Biochemistry* 36, 15315-15325.

Tayeb, H., Fiallo, M., & Garnier-Suillerot, A. (1996) *Chirality* 8, 585-589.

Tendler, M. D., & Korman, S. (1963) *Nature* 199, 501.

Teng, M.-K., Usman, N., Frederick, C. A., & Wang, A. H.-J. (1988) *Nucleic Acids Res.* 16, 2671-2690.

Thompson, A. S., & Hurley, L. H. (1995) *J. Mol. Biol. 252,* 86-101.

Thompson, A. S., Fan, J.-Y., Sun, D., Hansen, M., & Hurley, L. H. (1995) *Biochemistry* 34, 11005-11016.

Thurston, D. E. (1999) *Br. J. Canc.* 80 (supplement 1), 65-85.

Trotta, E., D'Ambrosio, E., Ravagnan, G., & Paci, M. (1996) *J. Biol. Chem.* 271, 44, 27608-27614.

Turner, P. R., & Denny, W. A. (1996) *Mutation Res.* 335, 141-169.

Van Rosmalen, A., Cullinane, C., Cutts, S. M., & Phillips, D. R. (1995) *Nucleic Acids Res.* 23, 1, 42-50.

van Tellingen, O. V., Nooijen, W. J., Schaaf, L. J., van der Valk, M., van Asperen, J., Henrar, R. E. C., & Beijnen, J. H. (1998) *Cancer Res.* 58, 2410-2416.

Verweij, J. (1996) *Cancer Chemother. Pharmacol.* 38, S3-S10.

Walker, G. S., Fagerness, P. E., Farley, K. A., & Mizsak, S. A. (1997) *J. Heterocyc. Chem.* 34, 295.

Wang, A. H.-J., Ughetto, G., Quigley, G., & Rich, A. (1987) *Biochemistry* 26, 1152-1163.

Wang, J. Y.-T., Chao, M., & Wang, H.-J. (1995) in *Anthracycline Antibiotics,* (Priebe W., ed) 168-182, American Chemical Society, Washington D.C.

Warpehoski, M. A., & Bradford, V. S. (1988) *Tetrahedron Lett.* 29, 2, 131-134.

Warpehoski, M. A., & Harper, D. E. (1995) *J. Am. Chem. Soc.* 117, 2951-2952.

Warpehoski, M. A., Gebhard, I., Kelly, R. C., Krueger, W. C., Li, L. H., McGovren, J. P., Prairie, M. D., Wicnienski, N., & Wierenga. W. (1988) *J. Med. Chem.* 31, 590-603.

Wartell, R. M., Larson, J. E., & Wells, R. D. (1974) *J. Biol. Chem.* 249, 6719-6731.

Weiland, K. L., & Dooley, T. P. (1991) *Biochemistry* 30, 7559-7565.

Weiss, M. A., Patel, D. J., Sauer, R. T., & Karplus, M. (1984) *Nucleic Acids Res.* 12, 9, 4035-4047.

Westendorf, J., Aydin, M., Groth, G., Weller, O., & Marquardt, H. (1989) *Cancer Res.* 49, 5262-5266.

Westhof, E., Rubin-Carrex, C., & Fritsch, V. (1995) in *Computer Modelling in Molecular Biology,* Goodfellow, J. M. (ed.), 1-7, VCH Verlagsgesellschaft, Weinheim (Federal Republic of Germany) and VCH Publishers, New York NY(USA).

White, S., Baird, E. E., & Dervan, P. B. (1997) *Chem. Biol.* 4, 569-578.

White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., & Dervan, P. B. (1998) *Nature* 391, 468-471.

Wierenga, W., Bhuyan, B. K., Kelly, R. C., Krueger, W. C., Li, L. H., McGovren, J. P., Swenson, D, H., & Warpehoski, M. A. (1986) *Adv. Enzyme Regul.* 25, 141-155.

Wijmenga, S. S., Mooren, M. M. W., & Hilbers, C. W. (1993) *NMR of Macromolecules A Practical Approach* 217-288, Roberts, G. C. K. (ed.) IRL Press, Oxford, UK.

Williams, D. H., & Fleming I. (1987) *Spectroscopic methods in organic chemistry,* Fourth Edn. McGraw-Hill Book Company (UK) Limited, Maidenhead, Berkshire, England.

Wilson, W. D., Tanious, F. A., Barton, H. J., Strekowski, L., Boykin, D. W., & Jones, R. L. (1989) *J. Am. Chem. Soc.* 111, 5008-5010.

Wilson, W. D., Tanius, F. A., Barton, H. J., Joens, R. L., Fox. K., Wydra, R. L., & Strekowski, L. (1990) *Biochemistry* 29, 8452-8461.

Wolff, I., Bench, K., Beijnen, J., Bruntsch, U., Cavalli, F., de Jong, J., Groot, Y., van Tellingen, O., Wanders, J., & Sessa, C. (1996) *Clin. Cancer Res.* 2, 1717-1723.

Wood, A. A., Nunn, C. M., Czarny, A., Boykin, D. W., & Neidle S., (1995) *Nucleic Acids Res.* 23, 18, 3678-3684.

Woynarowski, J. M., McHugh, M. M., Gawron, L. S., & Beerman, T. A. (1995) *Biochemistry* 34, 13042-13050.

Wütherich, K. (1986) *NMR of Proteins and Nucleic Acids,* John Wiley and Sons, New York, USA.

Wyatt, M. D., Lee, M., Garbiras, B. J., Souhami, R. L., & Hartley, J. A. (1995) *Biochemistry* 34, 13034-13041.

Yoon, J.-H., & Lee. C.-S. (1998) *Arch. Pharm. Res.* 21, 4, 385-390.

Yoshida, M., Banville, D. L., & Shafer, R. H. (1990) *Biochemistry* 29, 6585-6592.

Young, M. A., Jayaram, B., & Beveridge, D. L. (1997) *J. Am. Chem. Soc.* 119, 59-69.

Zakrzewska, K., Lavery, R., & Pullman, B. (1983) *Nucleic Acids Res.* 11, 24, 8825-8839.

Zimmer, C. H. (1975) *Prog. Nucl. Acid Res. Mol. Biol.* 15, 285-318.

Zimmer, C. H., Luck, G., & Nüske, R. (1980) *Nucleic Acid Res.* 8, 2999-3010.

Zimmer, C., & Wähnert, U. (1986) *Prog. Biophys. Molec. Biol.* 47, 31-112.